



**A METHOD TO PREDICT COMPRESSOR STALL IN THE TF34-100
TURBOFAN ENGINE UTILIZING REAL-TIME PERFORMANCE DATA**

THESIS
June 2015

Shuxiang 'Albert' Li, Civ, USAF
G Trevor Jones, Civ, USAF

AFIT-ENV-MS-15-J-036

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

**A METHOD TO PREDICT COMPRESSOR STALL IN THE TF34-100
TURBOFAN ENGINE UTILIZING REAL-TIME PERFORMANCE DATA**

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Systems Engineering

Shuxiang 'Albert' Li, BS

Civilian, USAF

G. Trevor Jones, BS, MBA

Civilian, USAF

June 2015

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**A METHOD TO PREDICT COMPRESSOR STALL IN THE TF34-100
TURBOFAN ENGINE UTILIZING REAL-TIME PERFORMANCE DATA**

Shuxiang ‘Albert’ Li, BS

Civilian, USAF

G. Trevor Jones, BS, MBA

Civilian, USAF

Committee Membership:

Maj. Jason Freels, PhD
Chair

Dr. David Jacques
Member

Maj. Vhance Valencia, PhD
Member

ABSTRACT

Air Force current operations continue to undergo significant changes compelled by decreasing fiscal appropriations, aging aircraft, and personnel drawdown. The Air Force must effectively improve current maintenance operations in part to deal with these challenges. This study will explore the area of the A-10 aircraft fleet's TF34-100 high-pass turbo-fan engine sensor data to seek its deterioration modelling and prognostics capability. In futurity this will allow for achievement of greater confidence in predicting the compressor stall which leads to engine performance deterioration and a costly repair in maintenance. By utilizing an innovative method to forecast the probability of compressor stall, according to individual engine sensor data which has recently become available, it will be possible to achieve significant benefits in both maintenance planning and mission scheduling (which will greatly reduce the associated costs of maintenance servicing).

Acknowledgments

We would like to express our sincere appreciation to our faculty advisor, Major Jason Freels, for his guidance and support throughout the course of this thesis effort. The insight and experience are certainly appreciated. We would, also, like to thank Mr. Jeff Hanson of Air Force Engineering and Technical Services (AFETS), Mr. Robert Bowels and Mr. Kenneth Hurst of Southwest Research Institute (SwRI) for their insight and tutorial assistance in developing this research.

Shuxiang ‘Albert’ Li

G. Trevor Jones

TABLE OF CONTENTS

	Page
Abstract	iv
Table of Contents	vi
List of Figures	viii
I. Introduction	1
General Issue	1
Problem Statement.....	3
Investigative Questions	9
Methodology.....	9
Assumptions/Limitations.....	12
Implications	12
Preview	13
II. Research and Development Process.....	14
Overview	14
Processing the RTED Data.....	15
Check the ACF and PACF of VG Data.....	23
Performance Summary for arima(p,0,q) over Various Combined Flights Data.....	35
Model Adequacy Check for arima(20,0,6) Model	42
Model Adequacy Check for Probability Model of Compressor Stall Fault Event.....	50
Relevant Research – Application of Weight Filter for Times Series Data.....	59
III. Methodology	69
Overview	69
Test Subjects.....	69

Summary.....	71
IV. Analysis and Results.....	72
Results of arima(p,0,q) Model.....	72
Results of Linear Regression Model (LRM).....	73
Investigative Questions Answered	78
V. Conclusions and Recommendations	79
Conclusions of Research	79
Significance of Research	79
Recommendations for Future Actions.....	79
Recommendations for Future Research.....	79
Summary.....	80
Bibliography	81
Appendix.....	82
Appendix A. The formula to calculate Variable Geometry (VG)	Error! Bookmark not defined.
Vita	83

LIST OF FIGURES

	Page
Figure 1: Locations illustrated for three interested sensors (T2C, NG and IGV).....	5
Figure 2: Physical location of the engine sensor Compressor Inlet Temperature (T2C)....	6
Figure 3: Physical location of engine sensor PLA.....	7
Figure 4: Physical location of the engine sensor IGV.	7
Figure 5: Physical location of the engine sensor PT5 and PS3.....	8
Figure 6: Physical location of the engine sensor for Front Frame Accelerometer.	8
Figure 7: Screenshot of ASIST opening RTED data file.....	15
Figure 8: Sample the RTED data files from all available flight data - Nellis hard disk and JRAMS website.....	16
Figure 9: Percentage of “good” data points over number of combined flights.	19
Figure 10: Statistical check for the mean of first order differenced Variable Geometry (VG) data. A range of amplitude 0.00002 to 0.00014 is observed from this plot, thus it is considered that all first order difference VG data is stationary.....	20
Figure 11: Time-series plots of multiple sensor data captured from the left and right engines of an aircraft 78-0657 during a single flight. (top row – T2C, middle row NG, bottom row – IGV). All sensor data points seem good.	21
Figure 12: Time-series plots of multiple sensor data captured from the left and right engines of aircraft 78-0657 during a single flight. (top row – T2C, middle row NG, bottom row – IGV) Left engine has “bad” T2C sensor data. In this kind of case, it is	

necessary to copy the good T2C data and apply it to the other engine because engines only physical located apart of 10.5 feet on aircraft.	22
Figure 13: One flight Variable Geometry (VG) data. Original sequence plot; First order differenced time series plot (stationarity for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.	23
Figure 14: Combined four flights Variable Geometry (VG) data. Original sequence plot; first order differenced time series plot (stationary for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.	24
Figure 15: Combined seven flights Variable Geometry (VG) data. Original sequence plot; first order differenced time series plot (stationary for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.	25
Figure 16: ACF plot for seven-combined-flights-data, which shows a quick decay pattern along the x-axis (number of lags).The blue dashed lines are 95% confidence interval.	26
Figure 17: PACF plot for seven-combined-flights-data, which indicates a not-so-quick decay pattern along the x-axis (number of lags). The blue dashed lines are 95% confidence interval.	27
Figure 18: Example of a sample data (four-combined-flights) to be used to explain the concept of Percentage of Significant Coefficients (PSC) of arima(p,0,q) model. ...	31
Figure 19: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for one flight Variable Geometry (VG) data. These plots are for <i>Sigma</i>² ; <i>Log – Likelihood</i> ; <i>AIC</i> and Percentage of Significant Coefficients (PSC).	32

Figure 20: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for four flights Variable Geometry (VG) data. These plots are for <i>Sigma^2; Log – Likelihood; AIC</i> and Percentage of Significant Coefficients (<i>PSC</i>).....	33
Figure 21: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for seven flights Variable Geometry (VG) data. These plots are for <i>Sigma^2; Log – Likelihood; AIC</i> and Percentage of Significant Coefficients (<i>PSC</i>).....	34
Figure 22: Example of the arima(14,0,4) model performance parameter (<i>sigma^2</i>) over various combined flight data.	36
Figure 23: Example of the arima(14,0,4) model performance parameter (<i>log – likelihood</i>) over various combined flight data.	37
Figure 24: Example of the arima(14,0,4) model performance parameter (<i>AIC</i>) over various combined flight data.	37
Figure 25: Example of the arima(14,0,4) model performance parameter Percentage of Significant Coefficients (<i>PSC</i>) over various combined flight data.....	38
Figure 26: Example of the arima(12,0,4) performance parameters (<i>sigma^2, log – likelihood, AIC</i> and Percentage of Significant Coefficients (<i>PSC</i>)) over each data.	39
Figure 27: Example of the arima(14,0,4) performance (<i>sigma^2, log – likelihood, AIC</i> and Percentage of Significant Coefficients (<i>PSC</i>)) over each data.....	40

Figure 28: Example of the arima(16,0,4) performance (<i>σ^2, log – likelihood, AIC</i> and Percentage of Significant Coefficients (<i>PSC</i>)) over each data.....	41
Figure 29: Model Adequacy check - one flight. The ARIMA(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).	44
Figure 30: Model Adequacy check - one flight. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.	44
Figure 31: Model Adequacy check - one flight. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().	45
Figure 32: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).	45
Figure 33: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.	46
Figure 34: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().	47
Figure 35: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).	47
Figure 36: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.	48

Figure 37: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().	49
Figure 38: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for one-flight-data with arima(12,0,4).	51
Figure 39: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the one-flight-data, the “best fit” is arima(12,0,4).	51
Figure 40: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for two-combined-flight-data with arima(16,0,6).	52
Figure 41: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the two-combined-flight-data, the “best fit” is arima(16,0,6).	52
Figure 42: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for three-combined-flight-data with arima(12,0,4).	53
Figure 43: The examination of Linear Regression Model (LRM) adequacy based on various arima(p,0,q) are shown here. For the three-combined-flight-data, the “best fit” is arima(12,0,4).	53
Figure 44: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for four-combined-flight-data with arima(14,0,6).	54

Figure 45: The examination of Linear Regression Model (LRM) adequacy based on various $\text{arima}(p,0,q)$ are shown here. For the four-combined-flight-data, the “best fit” is $\text{arima}(14,0,6)$	54
Figure 46: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for five-combined-flight-data with $\text{arima}(16,0,4)$	55
Figure 47: The examination of Linear Regression Model (LRM) adequacy based on various $\text{arima}(p,0,q)$ are shown here. For the five-combined-flight-data, the “best fit” is $\text{arima}(16,0,4)$	55
Figure 48: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for six-combined-flight-data with $\text{arima}(12,0,2)$	56
Figure 49: The examination of Linear Regression Model (LRM) adequacy based on various $\text{arima}(p,0,q)$ are shown here. For the six-combined-flight-data, the “best fit” is $\text{arima}(12,0,2)$	56
Figure 50: Check Linear Regression Model (LRM) adequacy with R built-in function <i>plot(lm{stats})</i> . This is an example for seven-combined-flight-data with $\text{arima}(16,0,2)$	57
Figure 51: The examination of Linear Regression Model (LRM) adequacy based on various $\text{arima}(p,0,q)$ are shown here. For the seven-combined-flight-data, the “best fit” is $\text{arima}(16,0,2)$	57

Figure 52: Engine Stall Probability Model Adequacy Check – <i>R2</i> and <i>Radj2</i> of “Best Fit” Models versus the number of combined flights.....	58
Figure 53: Shapes of the three filters experiments.....	61
Figure 54: The effect of Filter 15 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.	62
Figure 55: The effect of Filter 15 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.....	62
Figure 56: The effect of Filter 15. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of first order differenced Variable Geometry (VG) data.....	63
Figure 57: The effect of Filter 31 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.	64
Figure 58: The effect of Filter 31 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.....	65
Figure 59: The effect of Filter 31. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of the first order differenced Variable Geometry (VG) data.....	65
Figure 60: The effect of Filter 61 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.	67
Figure 61: The effect of Filter 61 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.....	67

Figure 62: The effect of Filter 61. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of the first order differenced Variable Geometry (VG) data.....	68
Figure 63: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(12,0,4) for one-Flight-Data.....	74
Figure 64: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(14,0,6) for four-Combined-Flight-Data.....	75
Figure 65: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(16,0,2) for seven-Combined-Flight-Data.	76

A METHOD TO PREDICT COMPRESSOR STALL IN THE TF34-100 TURBOFAN ENGINE UTILIZING REAL-TIME PERFORMANCE DATA

I. INTRODUCTION

General Issue

The aim of preventive maintenance is to reduce the number of unexpected downtimes and therefore, the number of unscheduled maintenance actions. Unscheduled maintenance is an undesirable situation in which a failed component must be repaired or replaced before the system can return to service. Such unprecedented failures can interrupt delivery schedules, cause further system damage, and result in additional monetary burdens. For many components, maintenance may best be carried out in a proactive and preventive manner. This research focuses on the application of preventive maintenance for the TF34-100 jet engine to prevent engine compressor stalls for the A-10 aircraft. Due to their destructive nature, compressor stalls are a significant concern in axial flow compressor jet engines.

A compressor stall is caused by air approaching the compressor blades at an angle greater than their stalling angle resulting in a localized disruption of the airflow. The compressor blades act like small, cambered wings with very high aspect ratios; like high aspect ratio wings, the blades have relatively low stalling angles. Below the stalling angle, an increase in the angle of attack produces a proportional increase in the coefficient of lift. But, at angles beyond the stalling angle, the airflow separates from the upper surfaces. This causes a rapid decrease in coefficient of lift, with the departing air forming a highly turbulent wake downstream of the blade. This turbulent flow then moves downstream passes over the stator blades behind and into the next row of rotors. The process from this point onwards can take a number of forms. In

many cases the turbulence is simply flushed through the engine, such that only one or two blades stall. In this case there will be few if any outward indications that a stall has occurred. In the next level of severity, the initial stall might cascade rearwards, such that several rows of blades are affected, before the subsequent stages regain control of the air. In this case, the pilot might detect a sudden increase in internal temperature and increased vibration from the engine. The next level of severity, with the stall cascading all the way back through the compressor, will obviously increase the severity of the outward symptoms. Some engine types can behave in quite bizarre manners. Some US military turbofans introduced in the seventies or eighties, for example, exhibited what is termed a locked-in rotating stall. In this case the stalling of a single blade cascaded not downstream, but onto the next rotor blade on that disc. This caused a small pocket of stall to rotate in the opposite direction to the engine, but at about half of the engine's revolutions per minute (RPM). The engine would continue to run and produce some thrust.

A blade's stalling angle is not entirely fixed. Structural damage caused by the ingestion of hard objects, sand, and de-icing fluid can all reduce the stalling angle. Engines tend to stall more easily as they age, or if their compressors become iced or dirty. The degree to which the engine is able to control its airflow increases from front to rear. This is because the first row of rotor blades must accept whatever airflow they meet, whereas the subsequent rows receive air from the blades ahead of them.

The most common method of stall prevention is the use of variable angle inlet guide vanes (VIGV). These are an additional row of non-rotating blades, immediately ahead of the first row of rotors. By changing the pitch angle of these vanes, the angle of the incoming airflow is varied to ensure that the angle of attack of the first row of rotors is always less than their stalling angle. This method is employed in the TF34-100 A-10 engines.

Problem Statement

Until recently, sufficient data has not been available to elicit a predictive maintenance program for the A-10. However, a recent modification to the A-10 data accumulation ability has made available a broad range of sensor data. The current A-10 compressor stall maintenance process is to apply corrective maintenance after a stall has occurred. This reactive process can produce many issues, not the least of which is safety to the pilot when a compressor stalls during flight. Developing a predictive preventative maintenance process for compressor stall, with the recent additional aforementioned engine sensor data, would greatly benefit the A-10 program.

A-10 engine SMEs that analyze compressor stall and other engine problems are using fault event statistical data, in an attempt to derive some form of a logistical strategy planning. Some engine studies examine the engine performance deterioration data to develop a type of preventive maintenance routine. (See articles listed in the biography.) There is an apparent lack of a predictive method for compressor stall (or similar problems) concerning each individual engine, which method would be based upon the associated individual engine's real time sensor data of that particular turbofan engine.

Research Flow

This study explores the relationship between real-time engine sensor data to engine compressor stalls to develop a quantifiable algorithm from this engine sensor data which can be applied toward predictive preventative maintenance.

The biggest initial question is if a connection between the compressor stall fault events to the particular parameters of engine performance sensor data can be derived. Numerous

discussions with engine subject matter experts (SMEs) and engine maintenance SMEs, concluded that compressor stall fault events would be most likely associated with the compressor discharge pressure (PS3), turbo discharge pressure (PT5), and variable geometry (VG).

Typical A-10 engine SMEs' primary investigations and discussions, suggests that the VG parameter is the leading indicator for the deterioration of the engine control mechanism which, in turn, is the major factor in the development of compressor stall. The VG measure was created by the engine manufacturer General Electric to assess the discrepancy between the theoretical value of the guide vane positioning, (the ideal value for an engine to perform under pristine working environmental conditions) and the actual value as measured. Current VG calculation algorithm does not show the "theoretical value" but only the difference between the theoretical value and the actual value.

When the A-10 engine and engine maintenance SMEs perform a compressor stall case study, they will discount any VG value out of ± 1.50 as a reference point to future problems. How the VG values associated to a compressor stall fault event and to what extent of its degree is unknown so far. Also it is needed to point out that, it is an experiment-derived formula for General Electric to calculate the VG value from three sensors. Those sensors are Compressor Inlet Temperature (T2C) sensor (also called CIT), the Core Speed sensor (NG), and the Inlet Guide Vane sensor (IGV).

The indication of this research is that the VG data contains information associated to the engine stall deterioration mechanism. By modeling this VG data of individual engines, the resulting modeling coefficients will be representatives of that deterioration mechanism. By collecting multiple flight data from multiple engines, the subsequent regression data model will

reveal the probability of a compressor stall event statistically related to such a deterioration mechanism.

The locations of certain engine sensors are shown in Figure 1: Locations illustrated for three interested sensors (T2C, NG and IGV).

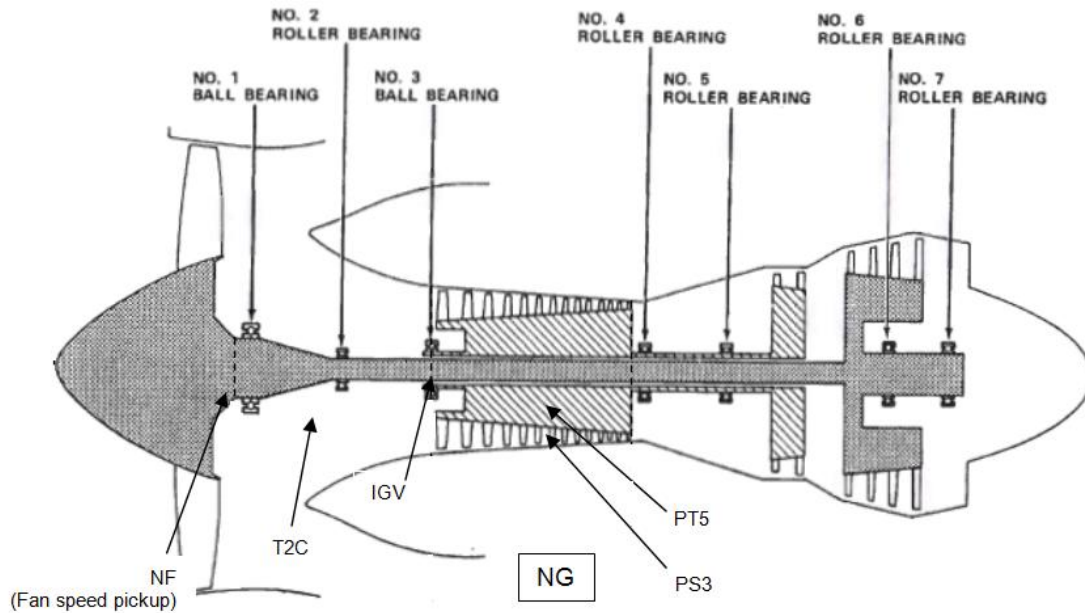


Figure 1: Locations illustrated for three interested sensors (T2C, NG and IGV).

The recent increase in available engine sensor data from the A-10 aircraft includes many different types of sensor data that vary in type and location. This research will particularly concentrate on compressor stall symptoms, which make up the majority cost of repair / maintenance for A-10 engines, by identifying a pattern of a precursor or constant in the now available recorded engine sensor data.

Two types of compressor stall are observed in practice where each type is characterized by a distinct set of sensor conditions.

Table 1: Types of compressor stalls

	WESA	AOA in Envelope
PLA	$\geq 12^\circ$	$\geq 12^\circ$
PS3 (% drop over 500ms)	$> 25\%$	$> 50\%$
NF	7%	
AOA	in envelope for mach number	in envelope for mach number

Photos illustrating the physical locations of several sensors are shown in Figures 3 – 6 below.

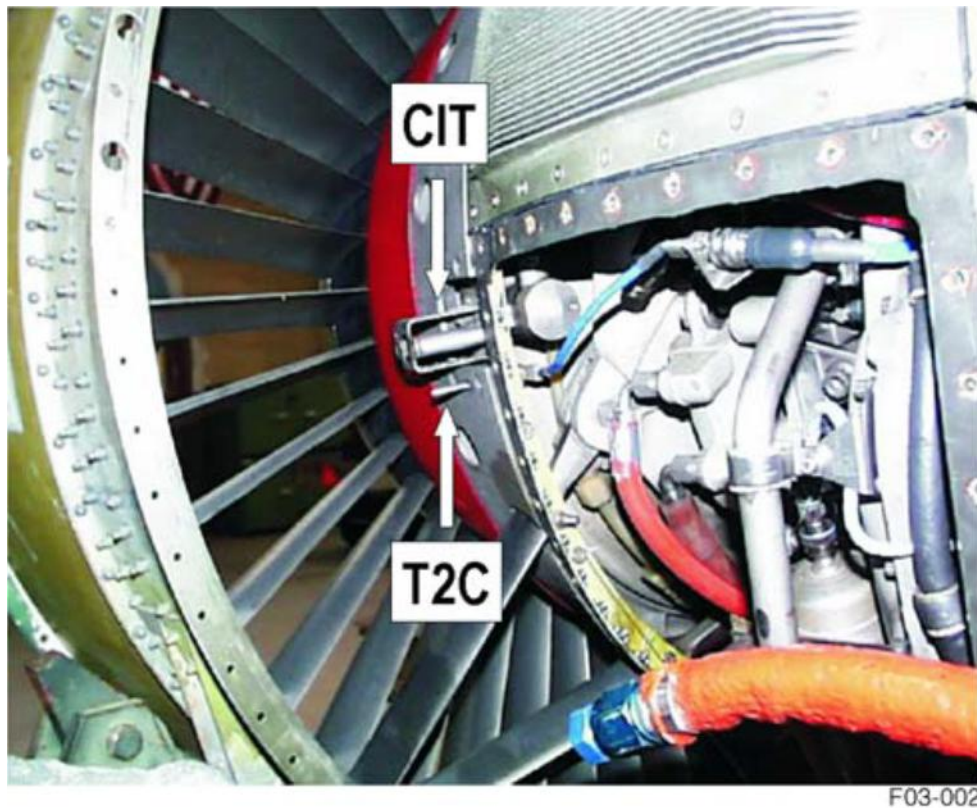


Figure 2: Physical location of the engine sensor Compressor Inlet Temperature (T2C).

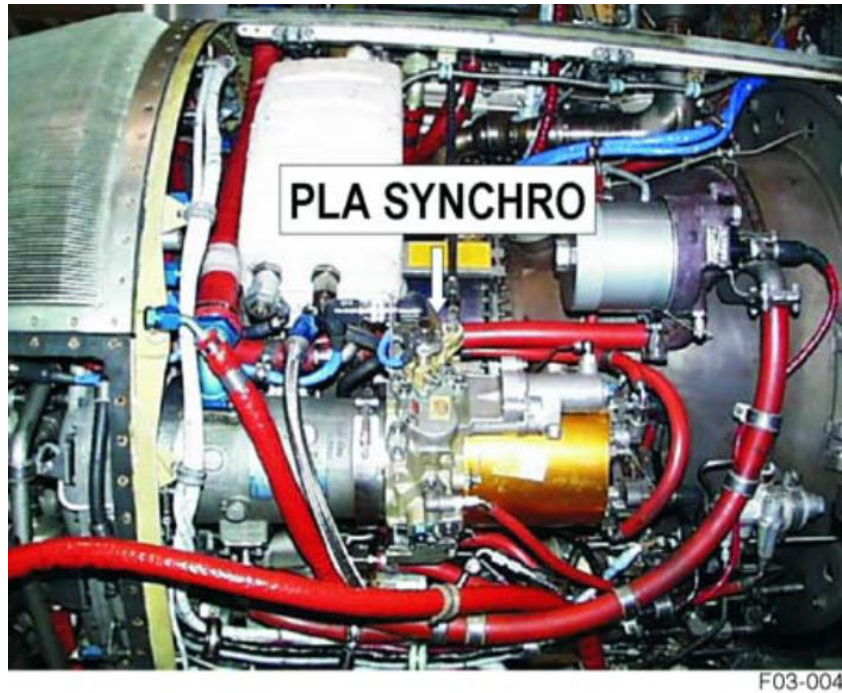


Figure 3: Physical location of engine sensor PLA.

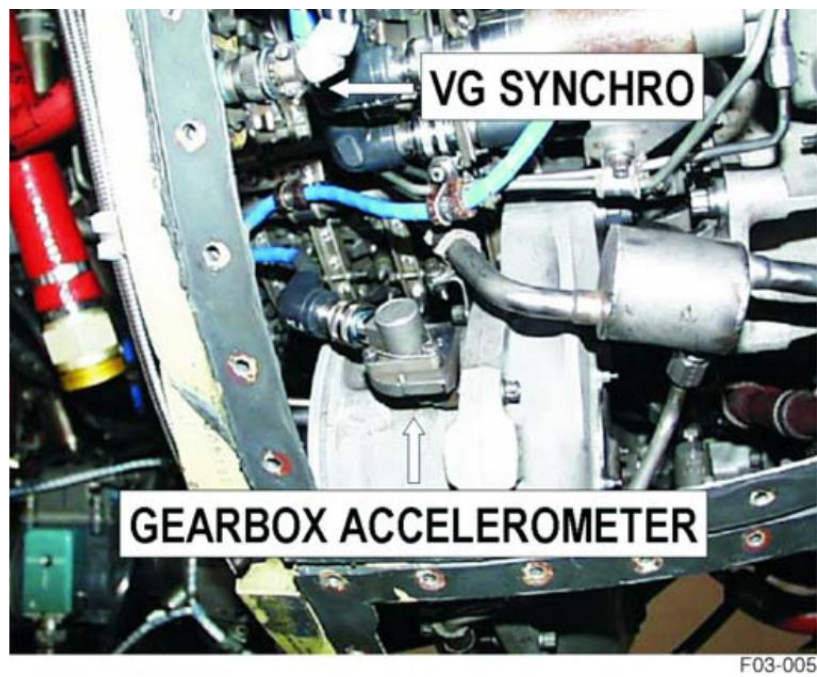


Figure 4: Physical location of the engine sensor IGV.

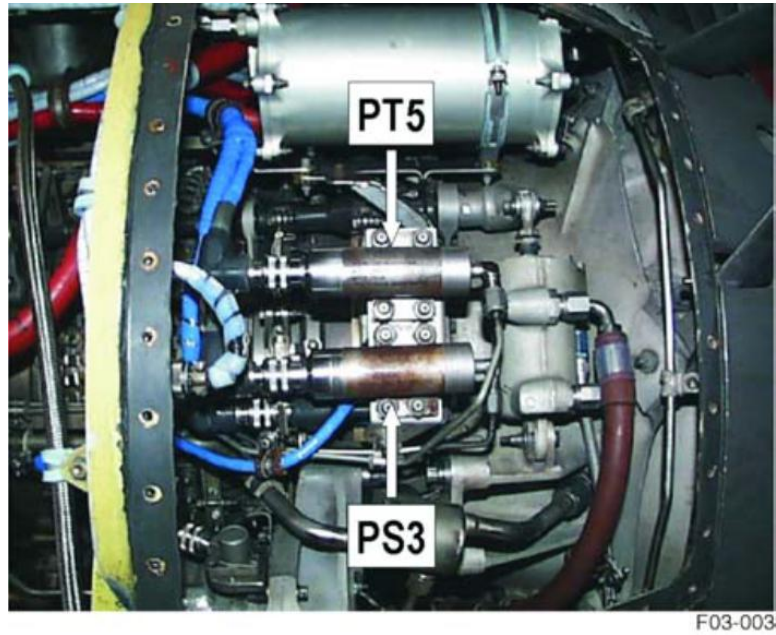


Figure 5: Physical location of the engine sensor PT5 and PS3.

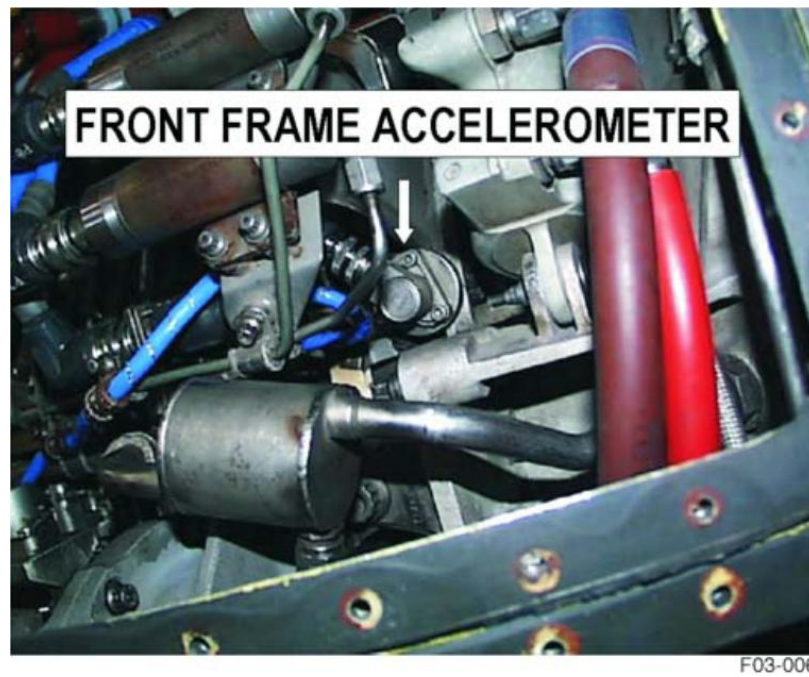


Figure 6: Physical location of the engine sensor for Front Frame Accelerometer.

Investigative Questions

How to sample the flight sensor data?

A large A-10 engine Real Time Engine Data (RTED) database is needed, one with consistency and reliability for modeling purposes.

How to derive a response indicator (independent variable) in order to establish a regression model?

There are many faults within the RTED data, how to identify the Compressor Stall indicator?

Is the AutoRegressive Integrated Moving Average (ARIMA) model adequate for the sensor data?

What order of the ARIMA model should be used? There is a need to understand which type of sensor data to use, one sensor or many sensors? In the case of many sensors, what transformation of the data is required to avoid a multivariate scenario?

How to predict the fault event probability by utilizing an ARIMA model?

Should the forecasting function of the ARIMA model be used to set an Upper Confidence Interval (UCL)/Lower Confidence Interval (LCL) band, to indicate the warning level of probability of future Compressor Stall fault event? Can a probability model be established by retrieving the Auto Regression (AR) and Moving Average (MA) coefficients from the selected ARIMA models?

Methodology

Engine sensor data is Time Series data: raw data that is collected from the General Electric TF34-100 engine sensors after each sortie, using a software program entitled “ASIST” developed by Southwest Research Institute (SwRI). This study will fit the raw flight sensor data

with an ARIMA model. Utilizing the outcomes from the modeling, a probability model will then be fitted in order to determine the chance of compressor stall occurring.

To ensure that a high fidelity state is maintained within the study, the impact of two known variance issues will be addressed and reduced. First is the sensor variance. This is an inherent variance associated with the nature of real world engineering data. Secondly is the modeling variance. This variance is related to the adequacy of the proposed model.

The following strategic steps were conducted during this study's research. Sample data was collected to conduct a primary investigation. Existing software analysis tool, entitled ASIST, is used to detect the fault events after each flight. A large amount of effort was employed to filter out or reduce the extent of sensor failure, sensor normal and abrupt variance. Through trial and error it was found that ARIMA modeling is appropriate for the VG data. This required many attempts to identify an adequate model and the necessary programming for the ARIMA model. These results yielded the equation:

$$X_t = \delta + AR_1X_{t-1} + AR_2X_{t-2} + \dots + AR_pX_{t-p} + A_t - MA_1A_{t-1} - MA_2A_{t-2} - \dots - MA_qA_{t-q}$$

Where X_t is the VG value at the time t , AR_p is the AutoRegression (AR) coefficient, and MA_q is the Moving Average (MA) coefficient. This model is denoted as arima(p,0,q).

Next a large population data was collected to develop the associated model. This was followed by identifying the model variables and selecting an ARIMA model. The estimate parameters were identified for first order differenced VG data. Then the model was checked for adequacy. Finally a probability model was fitted for compressor stall symptoms. The resulting equation was developed as:

$$p(EngineStall|N_{flights}) = \beta_0 + \sum_{i=1}^p \beta_i * AR_i + \sum_{j=1}^q \beta_{p+j} * MA_j + \epsilon$$

Where $N_{flights}$ indicates number of flights data used, currently, $N = 1, 2, \dots, 7$. AR_i is the AR coefficients ($i = 1, 2, \dots, p$) and MA_j is the MA coefficients ($j = 1, 2, \dots, q$).

This Probability Model of Compressor Stall Fault Event is based on the data as described below

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} \sim \begin{pmatrix} AR_{11} & AR_{21} & \dots & AR_{p1} & MA_{11} & MA_{21} & \dots & MA_{q1} \\ AR_{12} & AR_{22} & \dots & AR_{p2} & MA_{12} & MA_{22} & \dots & MA_{q2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ AR_{1n} & AR_{2n} & \dots & AR_{pn} & MA_{1n} & MA_{2n} & \dots & MA_{qn} \end{pmatrix}$$

Where the response variable (independent variable) (y_1, y_2, \dots, y_m) indicates compressor stall fault event did happened for engine #1 to engine #m, and (y_{m+1}, \dots, y_n) indicates No stall fault event happened for engine #(m+1) to engine #n; AR_{ik} ($i = 1, 2, \dots, p$; $k = 1, 2, \dots, n$) and MA_{jk} ($j = 1, 2, \dots, q$, $k = 1, 2, \dots, n$) are the regressor variables (explanatory variables) which are AR coefficients and MA coefficients of the fitted arima(p,0,q) model from each set of VG data. Currently $m = 6$ and $n = 28$. This reflects that the data consists of 14 aircraft (equipped with a total of 28 engines), and that 6 of the engines have compressor stall fault events.

Assumptions/Limitations

The first assumption is that the flight sensor data file is consistent, without missing one flight as a whole or any in-flight data points.

The second assumption is that the VG calculations are a correct discrepancy function of the theoretical value compared to actual value for the engine control mechanism.

For the third assumption, it is assumed that the major contributing factor to a compressor stall fault event is the variable geometry (VG) and that the other factors are minor.

The data for a particular engine has to be tracked every flight, which includes events of changing the engine into the studied aircraft. In other words, the engine number needs to be consistent for any study. Thankfully the ASIST software is very useful in identifying particular engine series numbers contained in the RTED data files. For the purpose of this study, those RTED data files have to be converted to csv files, and then introduced to the R programming of this research.

The model adequacy of the developed compressor stall Fault Event Probability Model can be checked by the theoretical calculations presented in this study, but the verification of the modelling adequacy check for the real world data is not validated yet. That is, the forecasted data from the proposed compressor stall Fault Event Probability Model will eventually need to go through a verification test for certain duration of time with real world data, such as 12 to 24 months, to check the accuracy of the fitted model with the A-10 fleet of engines.

Implications

If a more proactive method to conduct maintenance can be achieved for the A-10 TF34-100 engine, then a great amount of benefit will be realized not only in cost savings for engine

repairs, but also in aircraft/pilot safety and longer engine lifetimes. That is, this method can save many millions of dollars, increase the engine's reliability, and ensure a greater safety for A-10 pilots.

Preview

The Air Force current A-10 engine maintenance program for stalls is reactionary to physical manifestations. This paper will provide a more proactive modeling that can lead to a predictive preventative maintenance program. This involves the examination of a large body of A-10 engine data maintained by General Electric (GE). The resultant modeling of the data will be utilized to present a compelling option to achieve significant benefits in both maintenance planning and mission scheduling, which will reduce the associated costs of maintenance servicing as well as increasing the level of safety for the pilots of the A-10.

II. RESEARCH AND DEVELOPMENT PROCESS

Overview

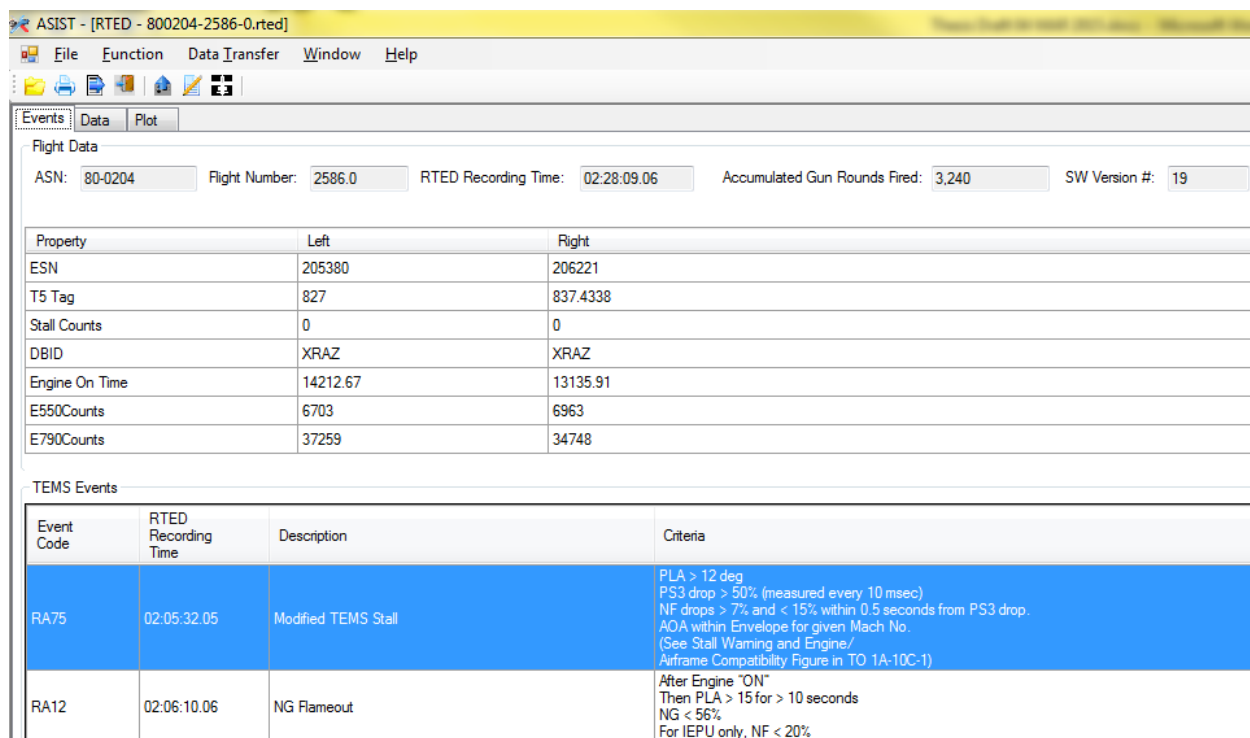
This chapter details the benefits of utilizing predictive preventative maintenance to enhance the current A-10 engine maintenance program. A brief background on the context areas is presented, followed by a more detailed review of literature pertaining to the process employed to develop the analysis. This will be the context used for the methodology and data analysis chapters. The intent is to present the reader a clear understanding of the statistical processes required to develop a model of the engine data.

The “micCID” is a legacy cartridge technology used by the U.S. Air Force to record real-time engine data (RTED) from various on board sensors. The data from each engine sensor is recorded, by the micCID, at a baseline rate of 1Hz, but if an abnormal behavior is detected this rate is increased to 100Hz. The data recorded at 100Hz are then saved as separate files from the data recorded at 1Hz. After each flight, the A-10 ASIST software is used to download the data from the micCID. Some RTED files are uploaded by A-10 aircraft deployed base technical personnel to a website called Joint Reliability Availability Management System (JRAMS). The RTED data used for this research were downloaded from JRAMS website and also delivered from a hard disk provided by Nellis AFB AFETS team.

Processing the RTED Data

Overview

The engine RTED file is downloaded from the aircraft, and ASIST analyzes the sensor data to find a fault event (engine stall and others). In particular, the ASIST tool is used to discover a fault event where the parameters match the definition of a compressor stall fault incident. For example, the screenshot in Figure 7 shows the fault event code “RA75”, indicating that while the aircraft was airborne, the right engine had an "Engine Stall with AoA in Envelope" event. Note, that this event is initially termed a “Modified TEMS Stall” in ASIST. This event was then assigned a numbered code of “75” in the engine SME’s archive. The ASIST software also records the engine serial numbers (ESN) as shown in Figure 7. Tracking the ESN is critical to this research to ensure that the same engine is selected for the seven flights of the analysis.



The screenshot displays the ASIST software interface with the title bar 'ASIST - [RTED - 800204-2586-0.rted]'. The menu bar includes 'File', 'Function', 'Data Transfer', 'Window', and 'Help'. Below the menu is a toolbar with icons for file operations. The main window is divided into two sections: 'Flight Data' and 'TEMS Events'.

Flight Data

Property	Left	Right
ASN	80-0204	
Flight Number	2586.0	
RTED Recording Time	02:28:09.06	
Accumulated Gun Rounds Fired	3,240	
SW Version #	19	
ESN	205380	206221
T5 Tag	827	837.4338
Stall Counts	0	0
DBID	XRAZ	XRAZ
Engine On Time	14212.67	13135.91
E550Counts	6703	6963
E790Counts	37259	34748

TEMS Events

Event Code	RTED Recording Time	Description	Criteria
RA75	02:05:32.05	Modified TEMS Stall	PLA > 12 deg PS3 drop > 50% (measured every 10 msec) NF drops > 7% and < 15% within 0.5 seconds from PS3 drop. AOA within Envelope for given Mach No. (See Stall Warning and Engine/Airframe Compatibility Figure in TO 1A-10C-1)
RA12	02:06:10.06	NG Flameout	After Engine "ON" Then PLA > 15 for > 10 seconds NG < 56% For IEPU only, NF < 20%

Figure 7: Screenshot of ASIST opening RTED data file.

The detailed data file sampling process is illustrated in Figure 8. This figure illustrates the possibility of multiple sets of flight data files to be selected from one particular aircraft, because some aircraft have more than one recorded compressor stall events. Since it is desired to obtain a broadband fleet representation, as much as possible, and in order to avoid repeatable errors from sample data (which would be the case if multiple flight data samples were used from the same aircraft), one set of sample data has been selected for each aircraft within this study.

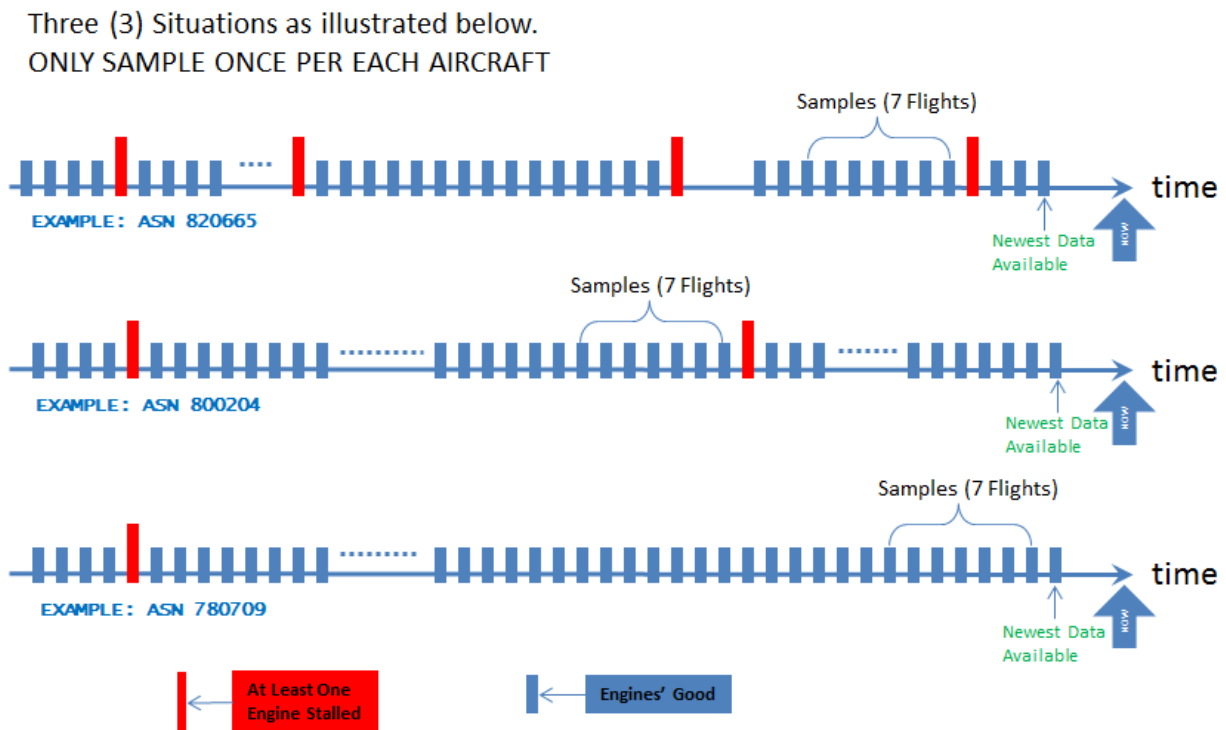


Figure 8: Sample the RTED data files from all available flight data - Nellis hard disk and JRAMS website.

Observe and choose all available RTED data for the continuous 7 flights of data as illustrated in Figure 8 above.

Use an R program to calculate the VG values and save the results into two separate data files for left-side and right-side engines respectively. Refer to existing GE documents for a definition of VG and its experimental meanings.

When checking the T2C sensor data, consider any data point with a value less than -15°C or larger than $+95^{\circ}\text{C}$ as a “bad” data point or “outlier”. The reasons for this is that: the engine SMEs advocate that any data point out of the range of -15°C to $+95^{\circ}\text{C}$ is not real, and that primary observations and investigations reveal that removing these particular T2C sensor data points leads to better ARIMA model. This rational is applied to both engine T2C sensors in order to eliminate bad T2C sensor data impact. If the count of “bad” data points from the right engine is larger than left engine, then the T2C sensor data from the left engine will be used instead. If the count of “bad” data points from the left engine is larger than right engine, then the T2C sensor data from right engine will be used instead.

Because the physical distance between the left engine air inlet and the right engine is approximately 10.5 feet, the left engine T2C sensor values will normally be equal or very close to the right engine T2C sensor. When “bad” T2C sensor data is detected, its entire T2C data will be eliminated as a whole, then the copy of the “good” data points from another engine are used for the calculations. (Please see Figure 12.)

When the calculated VG value is larger than 20 degrees, while it's $\text{T2C} > 23.8^{\circ}\text{C}$, a VG calculation formula with $\text{T2C} < 23.8^{\circ}\text{C}$ is applied as recommended by the Engine SMEs because the T2C delay has most probably occurred at that time. The original VG formula from GE contains four regions where VG is not defined. (Please reference **Error! Reference source not found.**) When the NGC value is larger than 56.1798 for its full range of T2C values, and when the NGC is larger than 93.3933 while T2C value is larger than 37.7°C , the VG value has not

been defined, and they are marked with -11, -12, -13 and -14 in the developed R scripts according to the region in which the data falls into. Those marked data points will eventually be considered as “outliers” and will be eliminated from further calculation in the modeling process. The impact of this undefined data point elimination and the percentage of data points kept in calculated process, is shown in Figure 9. The worst case scenario indicates that only 0.663% of the data points are eliminated and that the impact of these undefined data point elimination is minor.

VG time series data is generated by combining flights 1 (one) through 7 (seven). A first order differenced VG time series data was then used to develop an $\text{arima}(p,0,q)$ model in R.

A compressor stall fault event probability model is created by deriving the associated AR and MA coefficients, and the model adequacy of the linear regression model (LRM) is checked with well-known classical theoretical calculations. That is, an R build-in function *plot(lm{stats})* is used, which employs four calculations for plotting: first, Residuals versus Fitted Values; second, Normal Probability Plot of Standardized Residuals; third, Squared Root of Standardized Residuals versus Fitted Values; and four, Leverage versus Standardized Residuals.

For the ARIMA model, it is necessary to check for stationarity and seasonality of the data. Stationarity means that the data has a mean of zero (or very closed to zero). Seasonality means the data has periodic fluctuations. This research did not observe any seasonality for the data, thus seasonality was not considered in the ARIMA model. Also, the first order differenced VG data demonstrated a good stationary quality, therefore, the version of $\text{ARIMA}(p,0,q)$ used was without any seasonality feature. To mathematically check the stationarity of first order differenced VG data, a determination is made as shown in Figure 10. From this calculation, a range of amplitude 0.00002 to 0.00014 is observed for mean values from one flight data to

seven-combined-flights-data. This statistic calculation provides a confidence check that any mean value impacts to the final Linear Regression Model (LRM) will be minor; therefore our final Probability Model of Compressor Stall Fault Event does not include a mean value as a regressor.

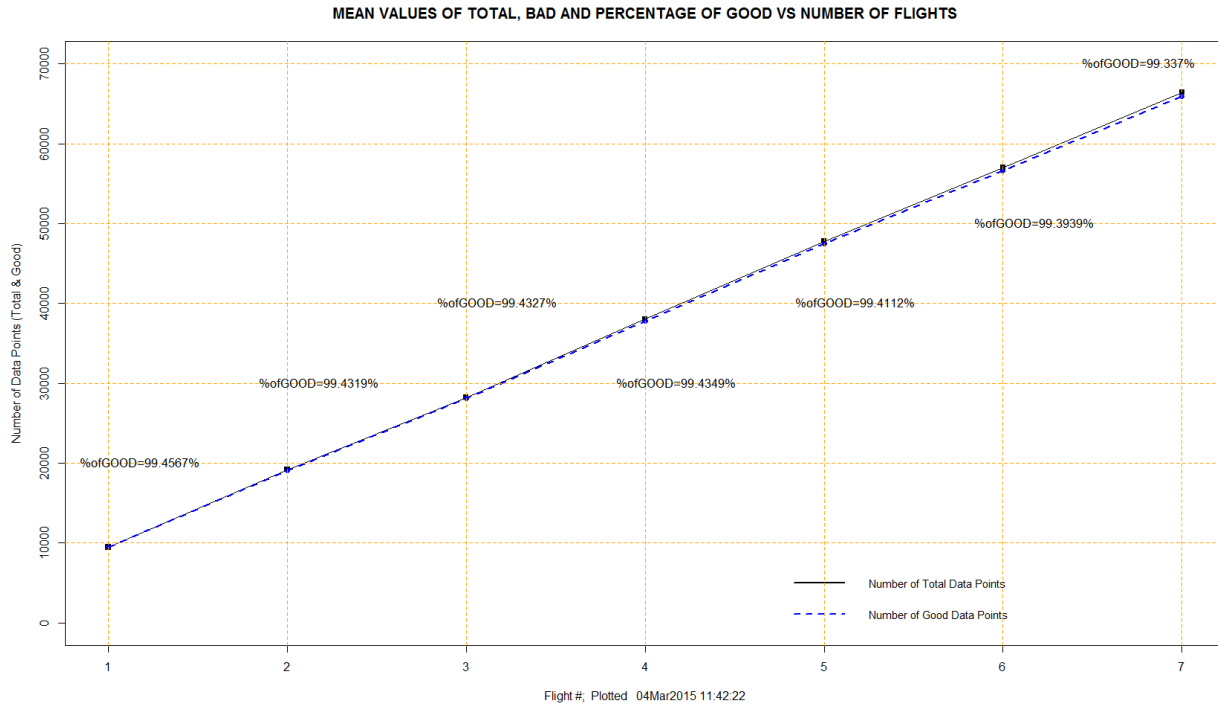


Figure 9: Percentage of “good” data points over number of combined flights.

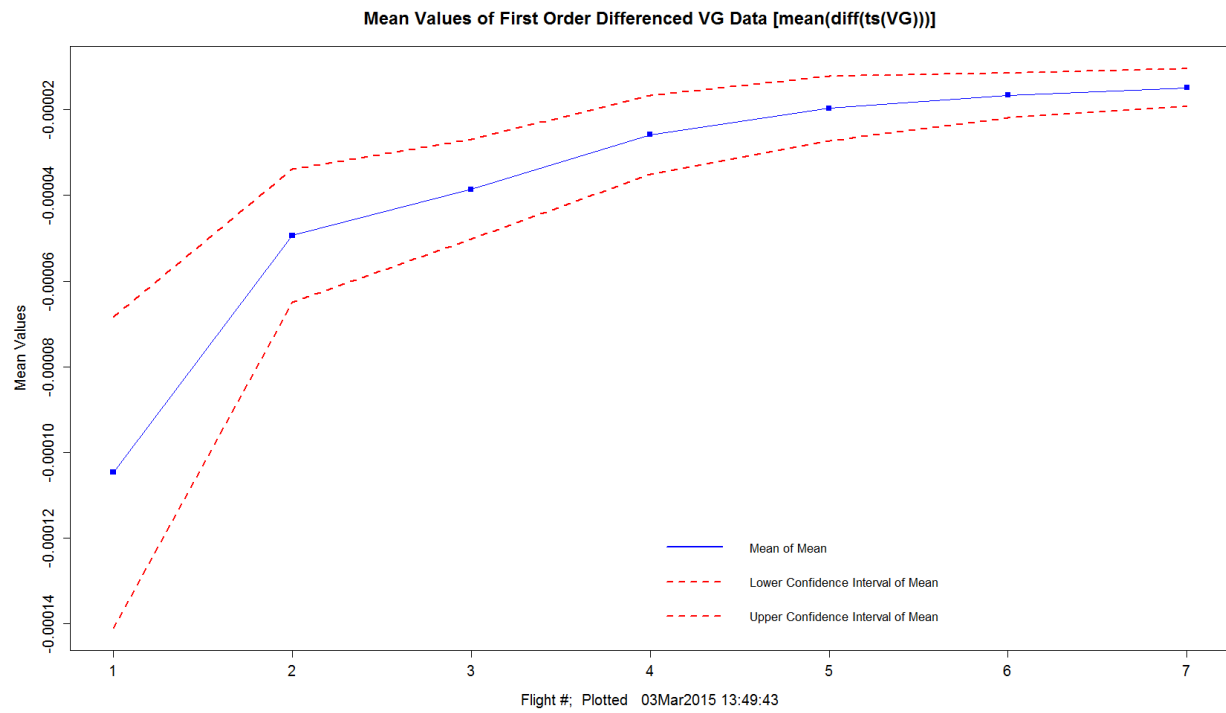


Figure 10: Statistical check for the mean of first order differenced Variable Geometry (VG) data. A range of amplitude 0.00002 to 0.00014 is observed from this plot, thus it is considered that all first order difference VG data is stationary.

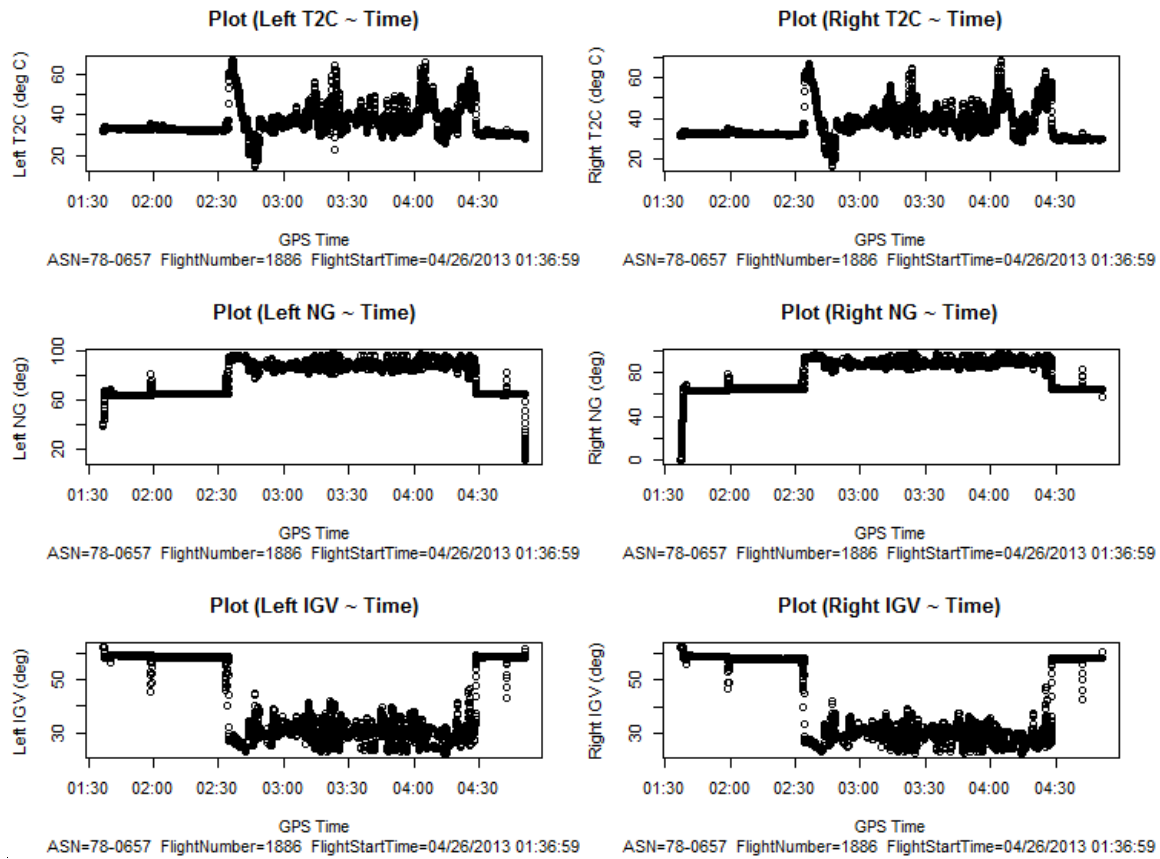


Figure 11: Time-series plots of multiple sensor data captured from the left and right engines of an aircraft 78-0657 during a single flight. (top row – T2C, middle row NG, bottom row – IGV). All sensor data points appear good.

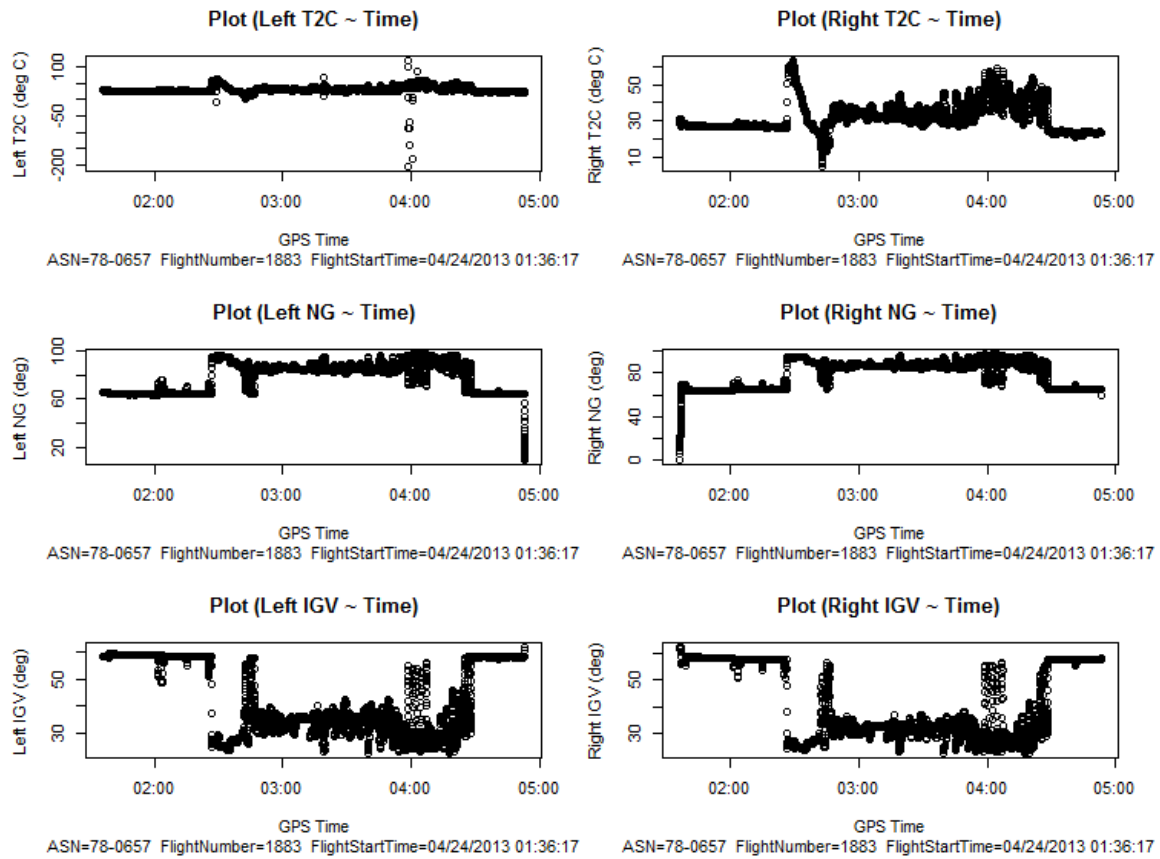


Figure 12: Time-series plots of multiple sensor data captured from the left and right engines of aircraft 78-0657 during a single flight. (top row – T2C, middle row NG, bottom row – IGV) Left engine has “bad” T2C sensor data. In this kind of case, it is necessary to copy the good (right engine) T2C data and apply it to this engine since the engines are physical located 10.5 feet apart on the aircraft.

Check the ACF and PACF of VG Data

The normal practice for ARIMA modeling is that, at first, by plotting autocorrelation function (ACF) can pre-determine a proper MA order of q in $ARIMA(p,0,q)$. Secondly, by plotting a partial autocorrelation function (PACF) can pre-determine a proper AR order of p in $ARIMA(p,0,q)$.

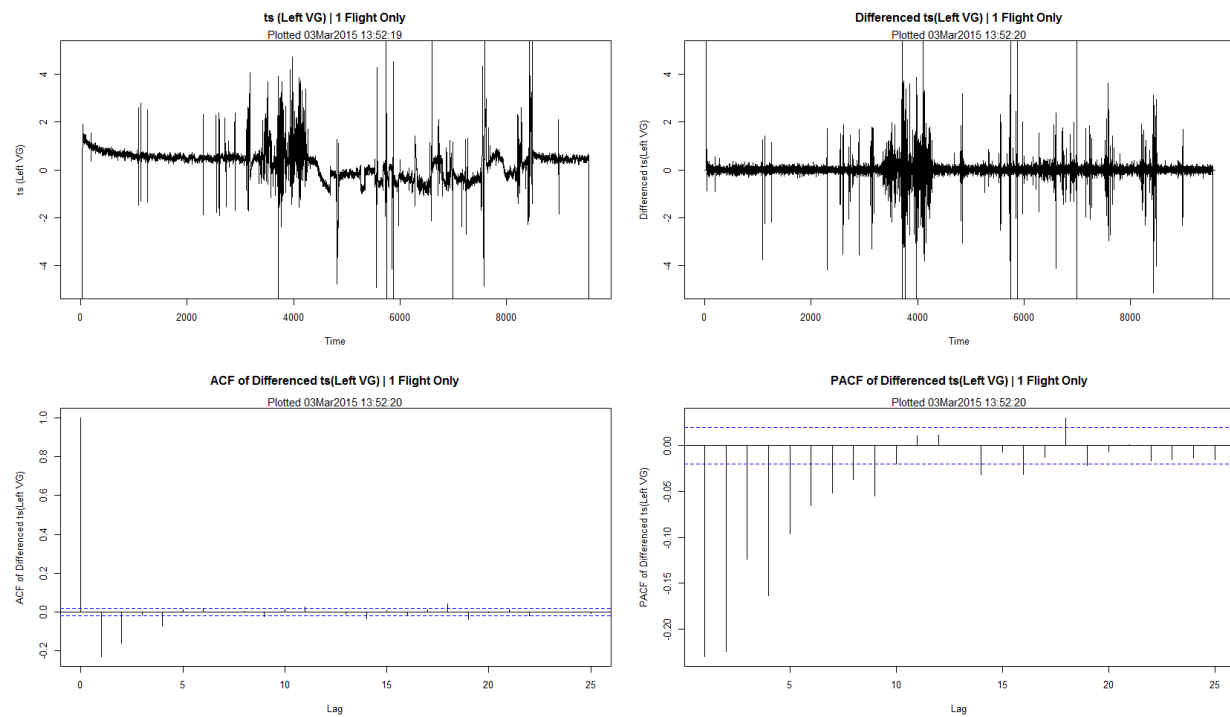


Figure 13: One flight Variable Geometry (VG) data. Original sequence plot; First order differenced time series plot (stationarity for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.

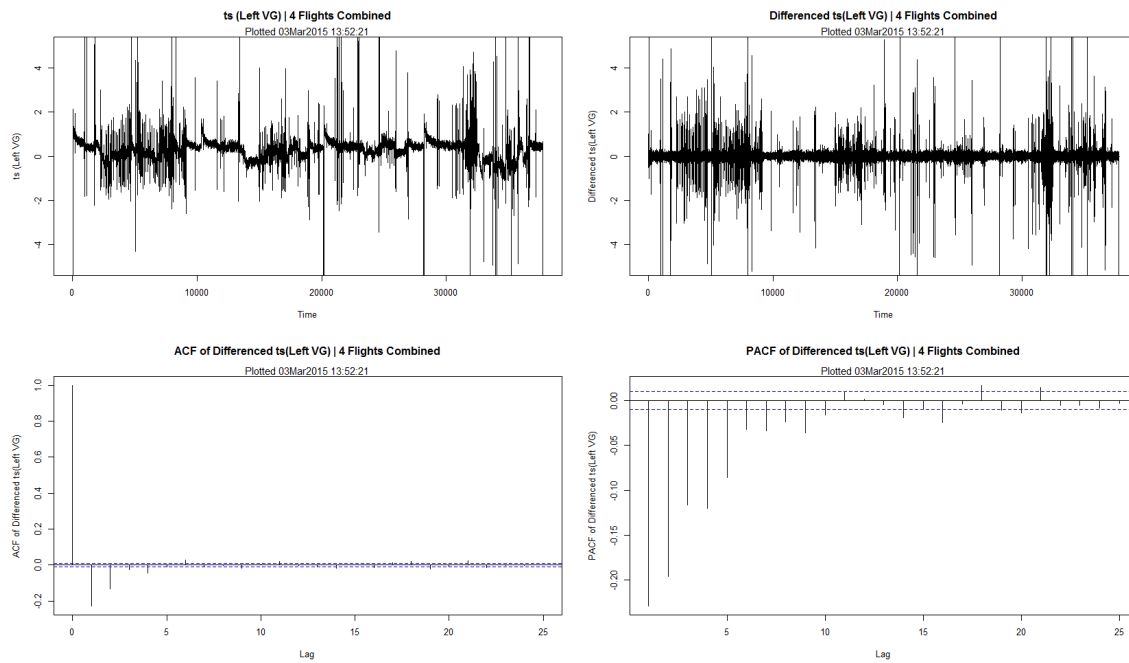


Figure 14: Combined four flights Variable Geometry (VG) data. Original sequence plot; first order differenced time series plot (stationary for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.

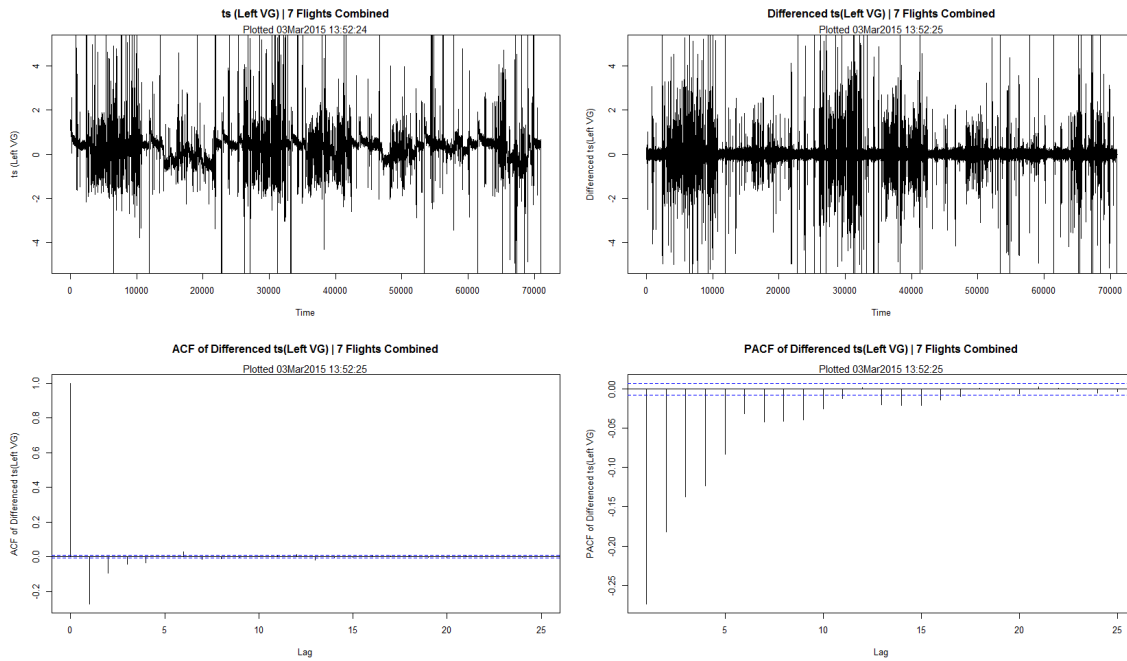


Figure 15: Combined seven flights Variable Geometry (VG) data. Original sequence plot; first order differenced time series plot (stationary for the ARIMA model); ACF plot and PACF plot for first order differenced ts VG data.

The ACF plot, such as Figure 16 shown below, shows a quick decay pattern along the x-axis (number of lags). After the lag = 6 point, all points are within a 95% confidence interval band. Especially after the lag = 2 points, the amplitude is very small (<0.1). This demonstrated plot suggests that having a value of $q = 6$ would be very adequate for the model. Even using a lower value of $q = 2$ may be adequate enough when using $\text{arima}(p,0,q)$. Similar ACF studies are carried out for many combined flight data. The findings are very similar to those indicated here.

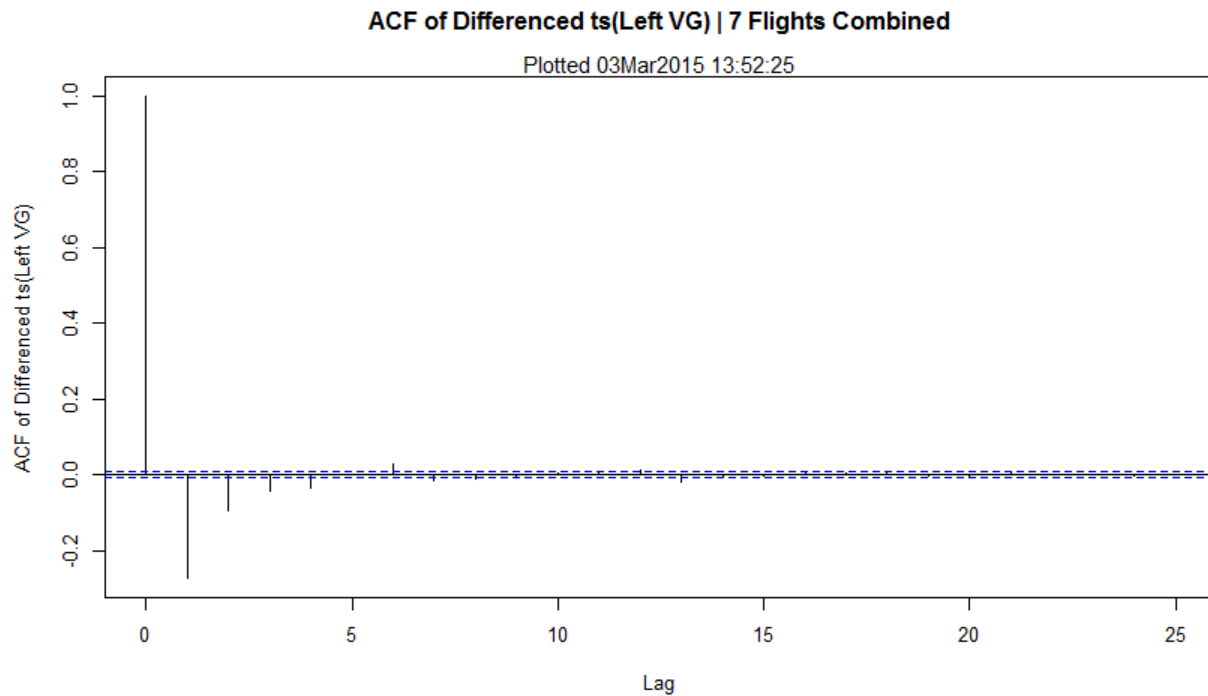


Figure 16: ACF plot for seven-combined-flights-data, which shows a quick decay pattern along the x-axis (number of lags). The blue dashed lines are 95% confidence interval.

The PACF plot, such as Figure 17 shown below, indicates a not-so-quick decay pattern along the x-axis (number of lags). After the lag = 20 point, all points are within a 95% confidence interval band. Especially after the lag = 16 points, the amplitude is very small (<0.025). This demonstrated plot suggests that setting $p = 20$ would be suitable for the model, and that setting $p = 16$ may be adequate enough when using $\text{arima}(p,0,q)$. Similar PCF studies are carried out for even more combined flight data. The findings are very similar as to those indicated here.

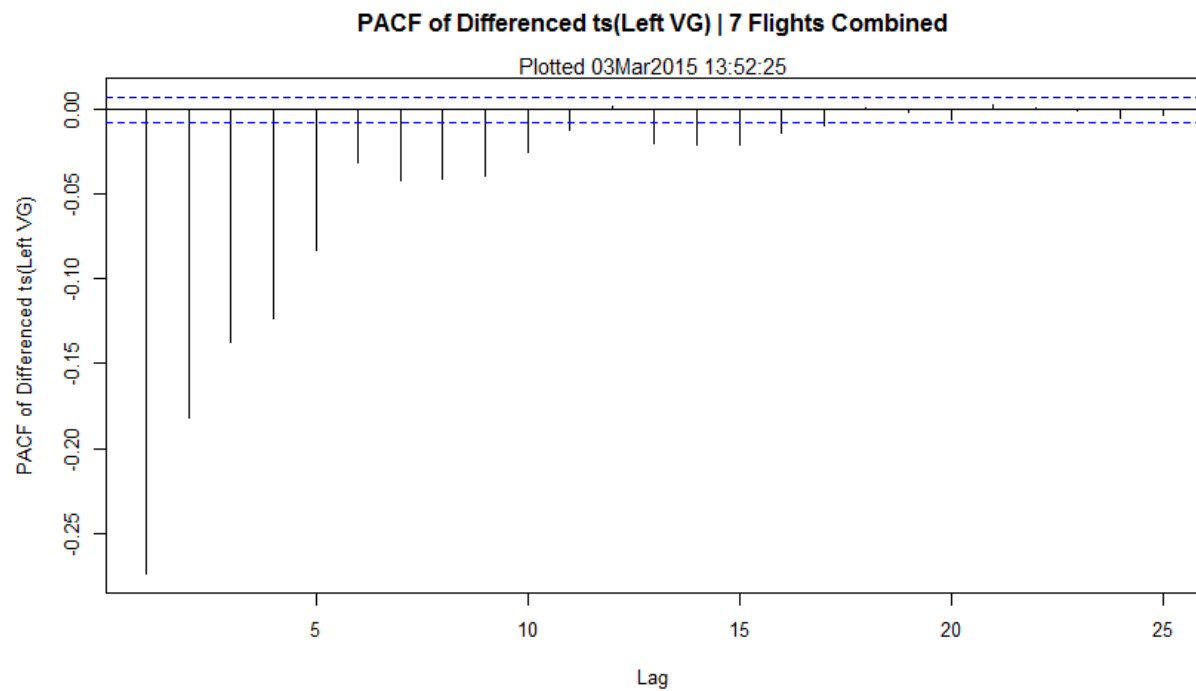


Figure 17: PACF plot for seven-combined-flights-data, which indicates a not-so-quick decay pattern along the x-axis (number of lags). The blue dashed lines are 95% confidence interval.

Check `arma(p,0,q)` Performance over Sample Data

A first order difference was formed from the VG data files for the sampled 28 engines, including the 6 engines with compressor stall faults using the flight data immediately prior to the fault flight for that particular aircraft. The aircraft serial number (ASN) and engine ID were checked for consistency to ensure that this was the case. Then 1 to 7 combined-flights-data were formed for the specified ASN aircraft. The reason to form a multiple flight-combined- data is to attempt to achieve an “early alert”.

An R script was developed to check a specified ARIMA(p,0,q) model's performance based on 4 (four) parameters.

First, check the Sigma^2 parameter by using the R code:

```
arma(x, order = c(p, 0, q), optim.method = "Nelder – Mead")$sigma2
```

Where Sigma^2 stands for the maximum likelihood estimate (MLE) of the innovations variance. The difference between the expected mean at time t, given the time series prior to t, and the actual value is called the innovation. Measuring the variance of the innovation will give you a better idea of how "noisy" the process is.

Second, check the *Log – Likelihood* parameter by using R code:

```
arma(x, order = c(p, 0, q), optim.method = "Nelder – Mead")$loglik
```

Where *Log – Likelihood* stands for a logarithm of likelihood function. In statistics, a Likelihood function (often simply the Likelihood) is a function of the parameters of a statistical model. The likelihood of a set of parameter values, θ , given outcomes x, is equal to the probability of those observed outcomes given those parameter values, that is:

$$\mathcal{L}(\theta|x) = P(x|\theta).$$

Third, check the *AIC* parameter by using R code:

```
arima(x, order = c(p, 0, q), optim.method = "Nelder – Mead")$aic
```

Where *AIC* stands for Akaike Information Criterion, which is a measure of the relative quality of a statistical model for a given set of data.

Fourth, check the Percentage of Significant Coefficients (*PSC*) parameter.

PSC represents the percentage of significant coefficients among the AR coefficients and MA coefficients of the fitted $\text{arima}(p,0,q)$ model. Let x be the first order differenced ts data of VG, $x.\text{fit}$ be the fitted $\text{arima}(p,0,q)$ model. That is:

```
x.fit <- arima(x, order = c(p, 0, q), optim.method = "Nelder – Mead")
```

If one of the coefficients of the fitted $\text{arima}(p,0,q)$ model $x.\text{fit}\$coef[k]$ ($k = 1, 2, \dots (p+q)$) meet following condition, it is considered as significant.

$$\left| \frac{x.\text{fit}\$coef[k]}{x.\text{fit}\$var.coef[k]} \right| > 1.96$$

In R programming environment, $x.\text{fit}\$coef[k]$ is the coefficients of the resulting $\text{arima}(p,0,q)$ model, where $x.\text{fit}\$coef[k]$ be AR coefficients while $k = 1, 2, \dots p$, and $x.\text{fit}\$coef[k]$ be MA coefficients while $k = (p+1), (p+2), \dots (p+q)$. Correspondingly, $x.\text{fit}\$var.coef[k]$ be the estimated variance of coefficient $x.\text{fit}\$coef[k]$.

Therefore, the *PSC* is calculated as:

$$PSC = \frac{\text{Sum of the Number of Significant } x.\text{fit}\$coef[k]}{(p + q)}$$

The following is a demonstration of the parameter's performance for the various ARIMA models and the different sampling data:

For example, process the sample data of Figure 18, is fit to an arima(12,0,4) model for the first order of VG data, by use of an R script, the *coefficients* of arima(12,0,4) is calculated as (-0.1815, 0.0002387, 0.01500, 0.03247, 0.02954, -0.002241, 0.05759, 0.04050, 0.006447, 0.01128, 0.01244, 0.02152, -0.05026, 0.04650, 0.02096, -0.01320)

Where the first 12 coefficients are called AR coefficients, and the last 4 coefficients are called MA coefficients, because $p = 12$ and $q = 4$ in an arima($p,0,q$) model here.

And the *variance of the coefficients* of arima(12,0,4) is calculated as (0.0003720, -0.0002028, -0.00002361, -0.0002780, 0.00001439, 0.00003668, 0.00005502, 0.00005025, 0.00005391, 0.00006573, 0.00003563, 0.00005364, -0.0003380, 0.0003338, -0.00001389, 0.0003176)

Thus, the ratio of coefficient to variance of coefficient $\left\{ \frac{\text{coefficient}}{\text{variance of coefficient}} \right\}$ is calculated as (487.84, 1.18, 635.56, 116.78, 2053.07, 61.09, 1046.79, 805.91, 119.60, 171.66, 349.30, 401.21, 148.69, 139.29, 1508.98, 41.57)

The second AR coefficient (AR_2) is not significant because its ratio of coefficient to variance of coefficient valued as of 1.18. The rest of the coefficients of arima(12,0,4) model are significant.

Therefore the *PSC* can be calculated as

$$PSC = \frac{\text{Sum of the Number of Significant } x.\text{fit\$coef}[k]}{(p + q)} = \frac{15}{16} = 0.9375 = 93.75\%$$

Lower than 100% indicating some coefficients are not significant. It is estimated that compounding factors and non-flight-state-separated data play roles to make some of the arima model coefficients non-significant.

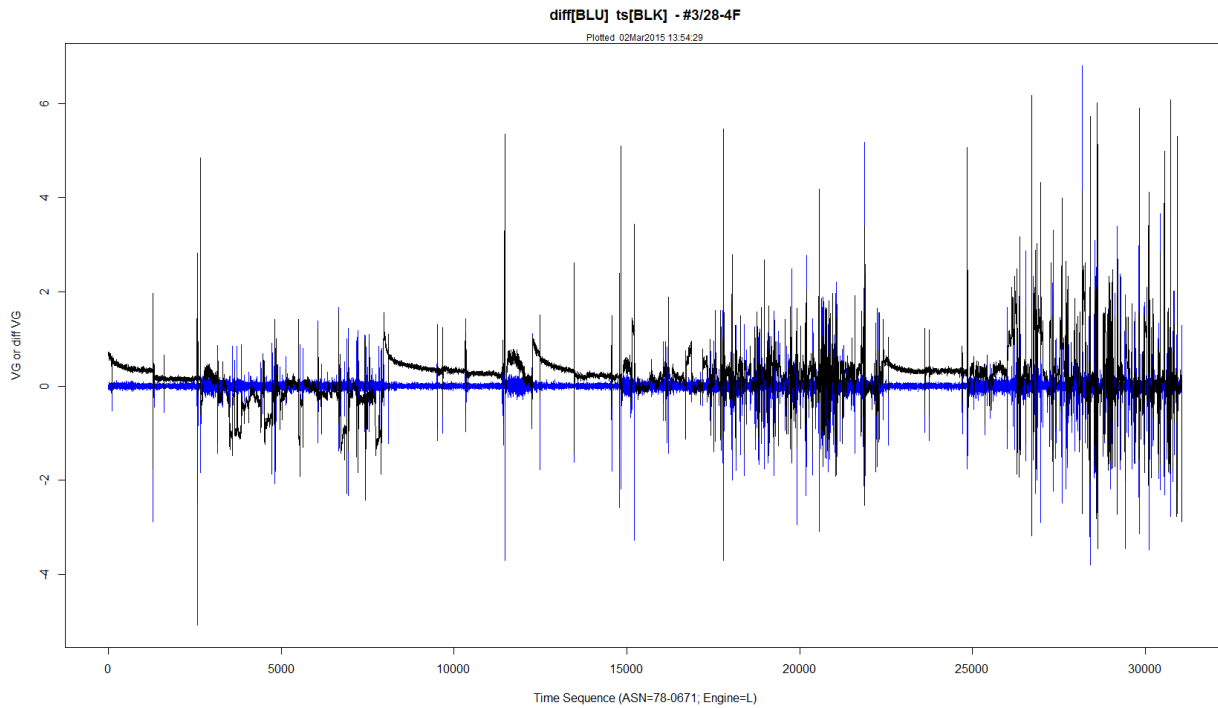


Figure 18: Example of a sample data (four-combined-flights) to be used to explain the concept of Percentage of Significant Coefficients (PSC) of $arma(p,0,q)$ model.

The following plots demonstrate the scan study results for various arima(p,0,q) model applied to 1-flight-data, 4-combined-flights-data and 7-combined-flights-data.

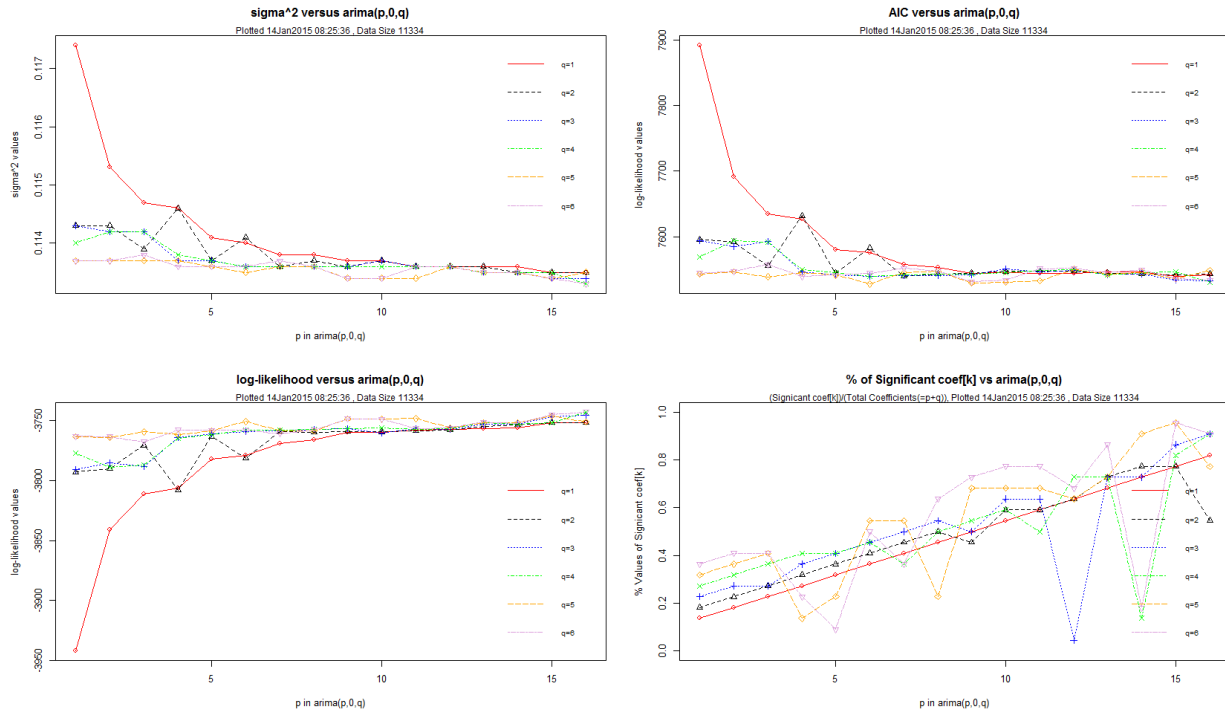


Figure 19: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for one flight Variable Geometry (VG) data. These plots are for σ^2 ; Log – Likelihood; AIC and Percentage of Significant Coefficients (PSC).

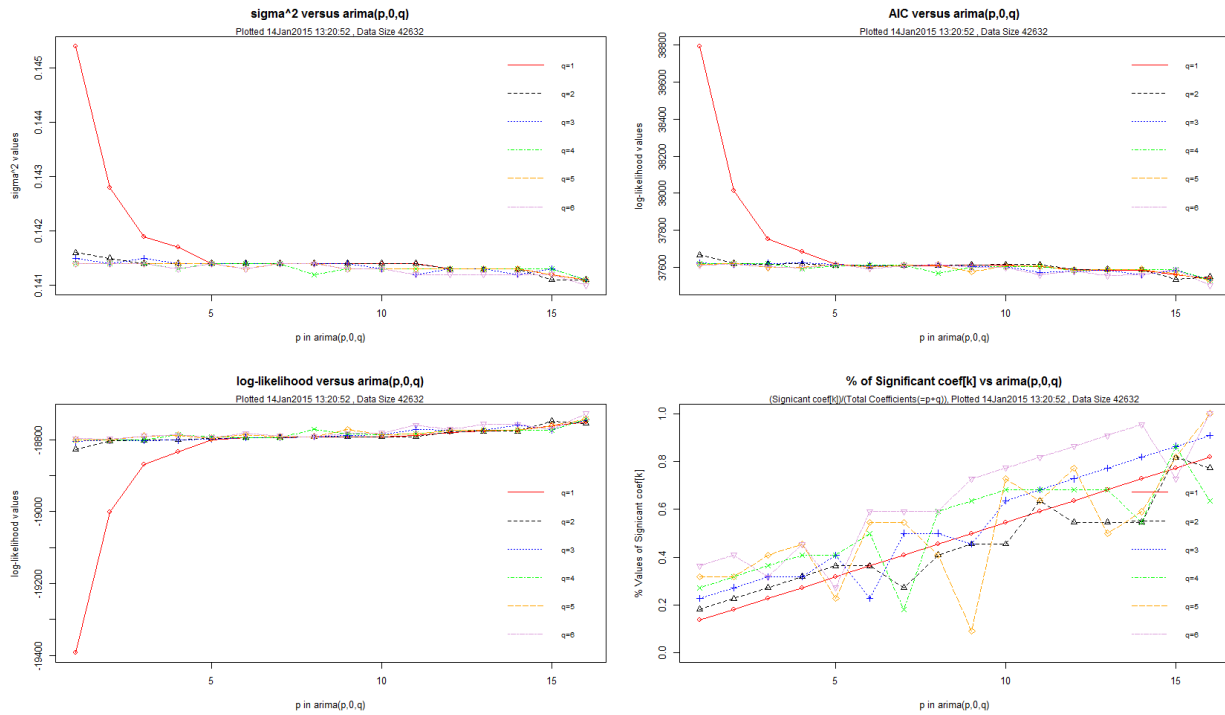


Figure 20: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for four flights Variable Geometry (VG) data. These plots are for σ^2 ; Log – Likelihood; AIC and Percentage of Significant Coefficients (PSC).

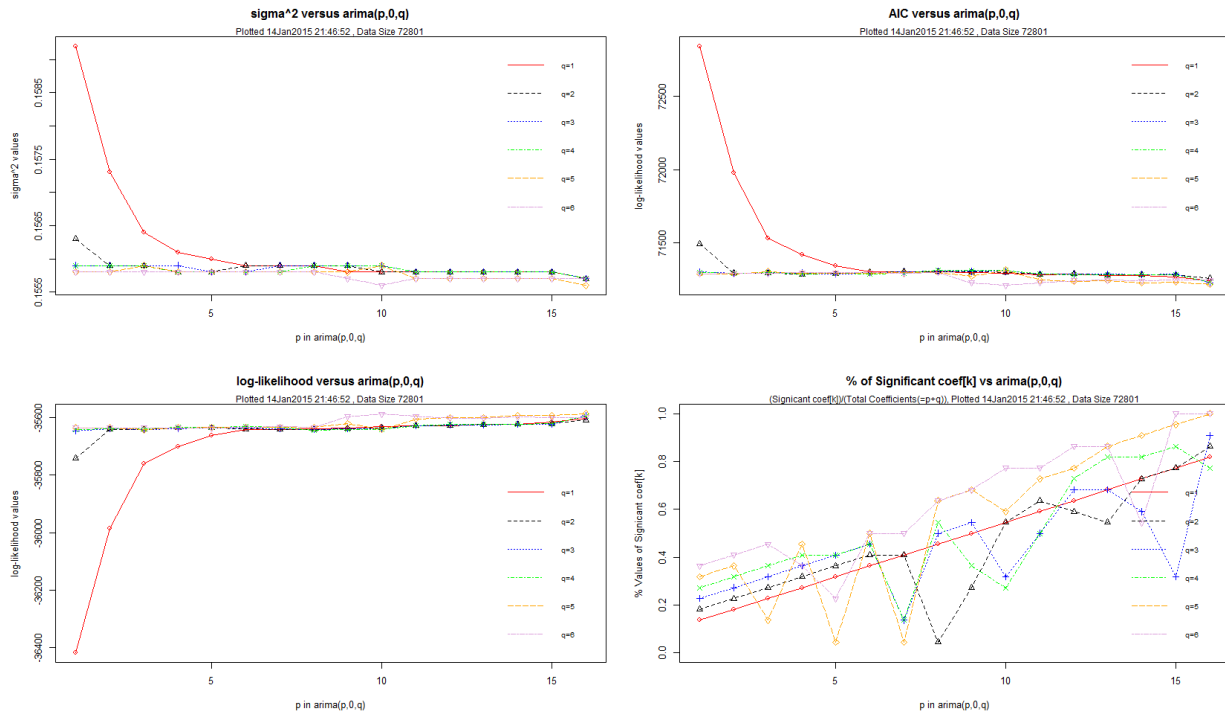


Figure 21: The four plots combination of the arima (p,0,q) model where p=1-16 and q=1-6. This is for seven flights Variable Geometry (VG) data. These plots are for σ^2 ; Log – Likelihood; AIC and Percentage of Significant Coefficients (PSC).

Performance Summary for arima(p,0,q) over Various Combined Flights Data

This study uses 28 sample engines, and have 1-flight-data up to 7-combined-flight-data, for a total of 196 (7x 28) samples. It was desired to determine how the arima(p,0,q) performed over all the data by experimenting with 9 arima(p,0,q) models, which included a combination of $p = 2, 4, 6$ and $q = 12, 14, 16$ in arima(p,0,q). All csv results and all plots have been retained for future reference and potential revalidation. A total of 22 samples with Negative fault (means no compressor stall is observed for next immediate flight) and 6 samples with Positive fault (means at least one compressor stall is observed for next immediate flight) have been examined respectively for their: mean values, UCI (Upper Confidence Interval) values, and LCI (Lower Confidence Interval) per 4 (four) performance parameters mentioned as above (σ^2 , $\log - likelihood$, AIC and PSC), versus 1-flight-data to 7-combined-flight-data.

The legends used are: Black ~ Negative, Red ~ Positive, Solid Lines ~ Mean, Dashed Lines ~ UCI, and Dotted Lines ~ LCI.

In general, the confidence interval can be calculated as

$$Confidence\ Interval = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Where x are the samples, \bar{x} is the mean of the samples, $(1-\alpha)$ is the confidence level, and $Z_{\alpha/2}$ is the confidence coefficient, σ is the standard deviation of the samples, and n is the sample size. Accordingly, the following R codes have been used to calculate the UCI and LCI with 95% confidence level

$$UCI <- -mean + qnorm(.025) * sd / sqrt(Sample\ Size\ of\ N\ or\ P)$$

$$LCI <- -mean - qnorm(.025) * sd / sqrt(Sample\ Size\ of\ N\ or\ P)$$

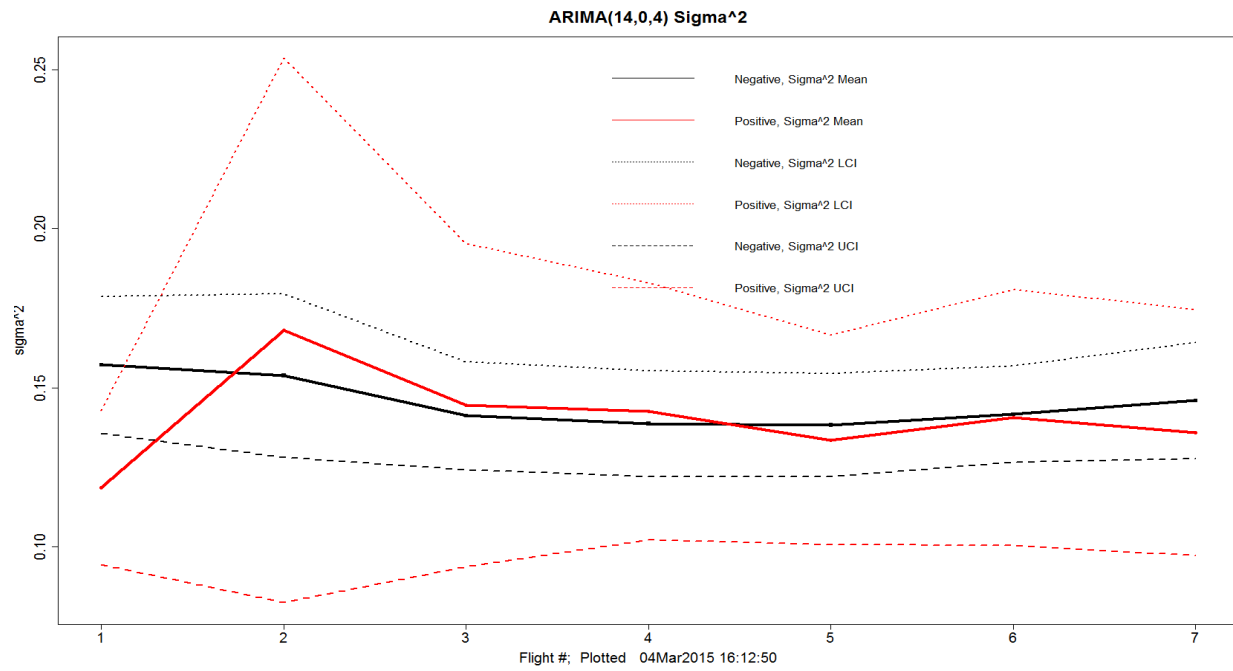


Figure 22: Example of the arima(14,0,4) model performance parameter (σ^2) over various combined flight data.

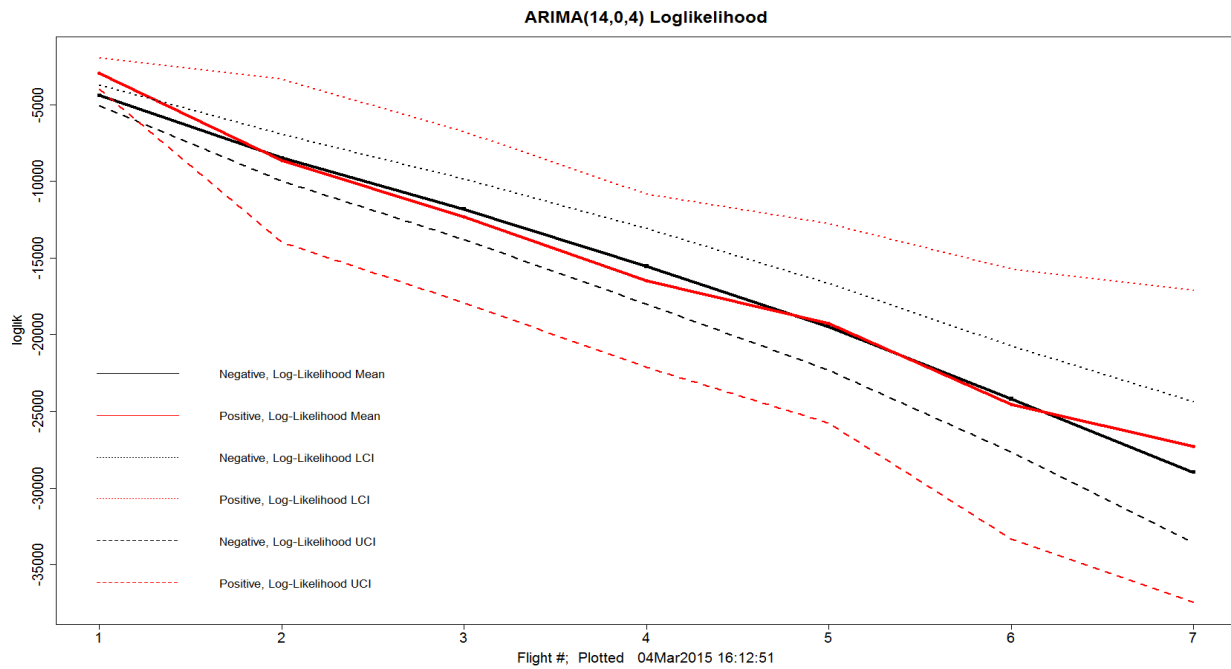


Figure 23: Example of the arima(14,0,4) model performance parameter (*log – likelihood*) over various combined flight data.

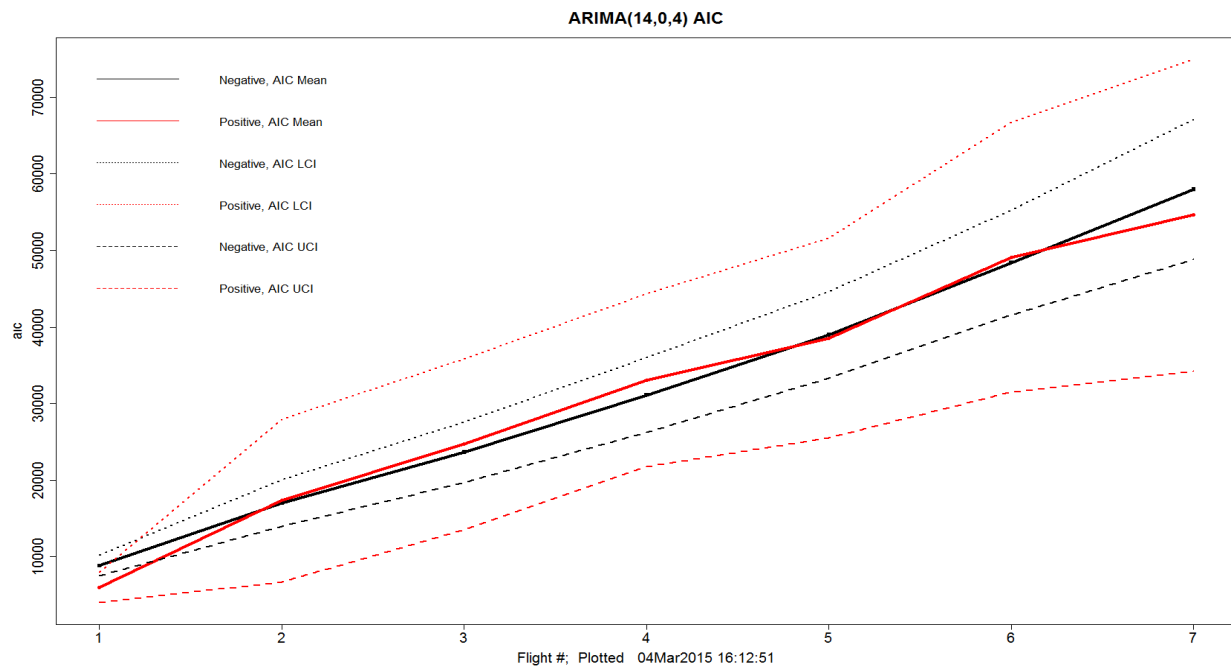


Figure 24: Example of the arima(14,0,4) model performance parameter (*AIC*) over various combined flight data.

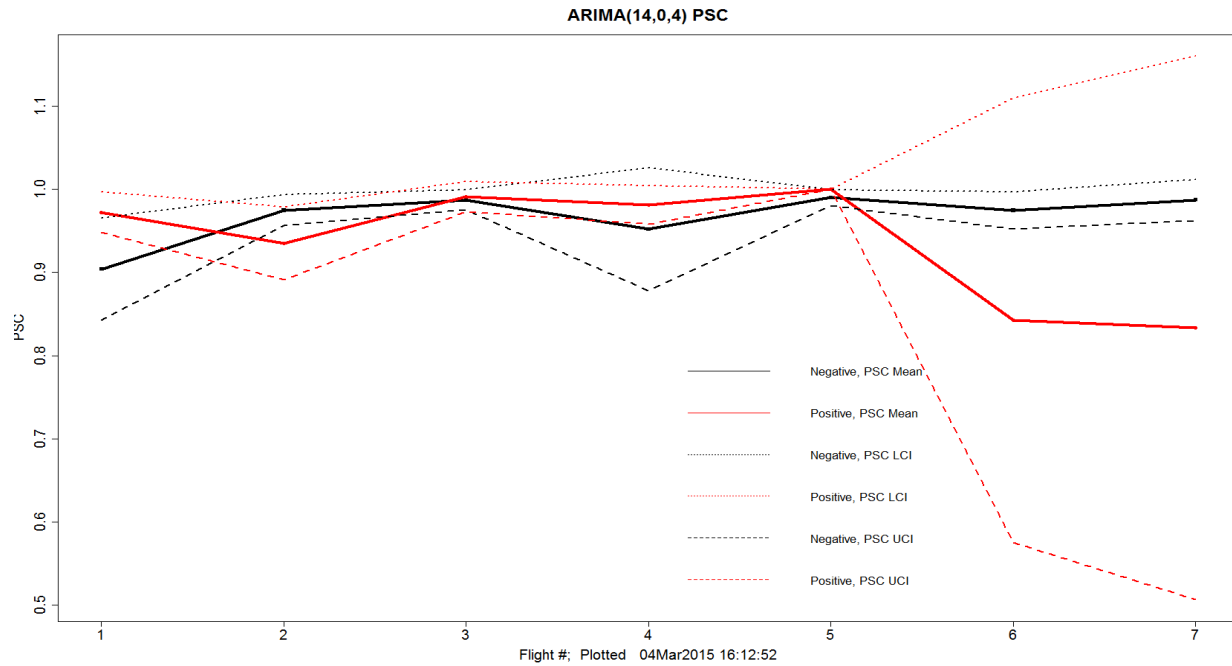


Figure 25: Example of the arima(14,0,4) model performance parameter Percentage of Significant Coefficients (*PSC*) over various combined flight data.

The following are the combined charts for various arima(p,0,q) performance parameters runs.

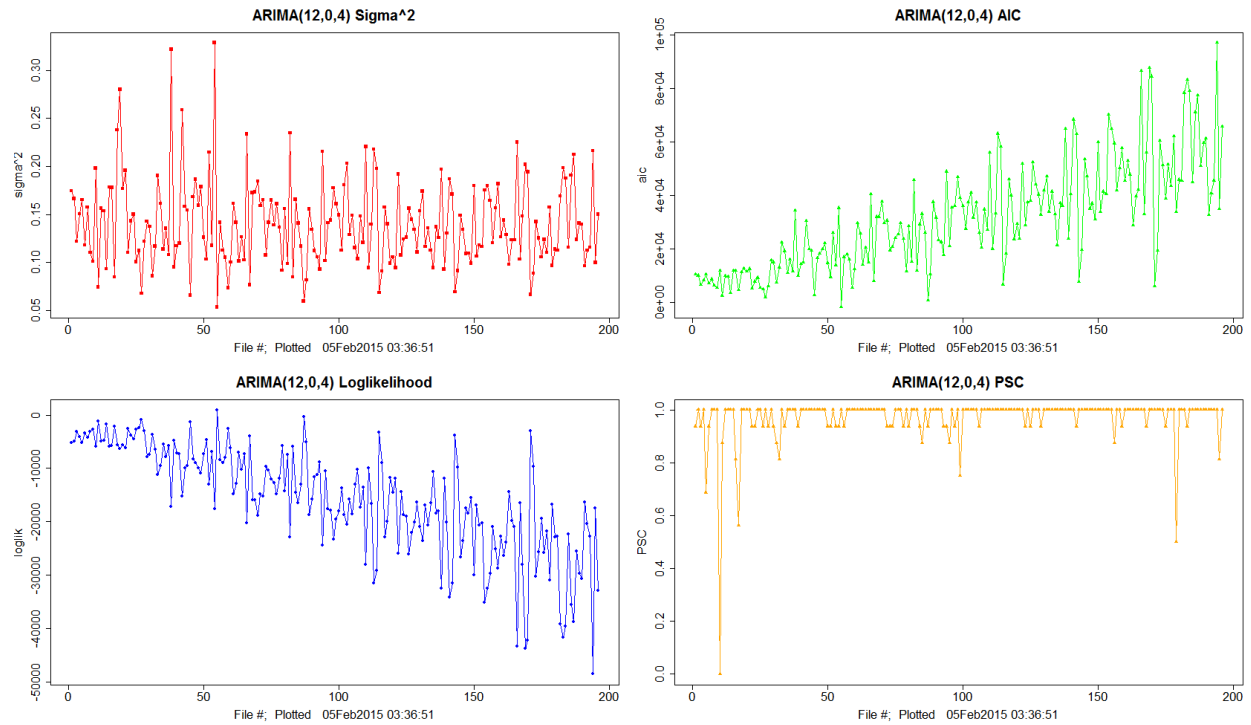


Figure 26: Example of the arima(12,0,4) performance parameters (σ^2 , $\log - \text{likelihood}$, AIC and Percentage of Significant Coefficients (PSC)) over each data.

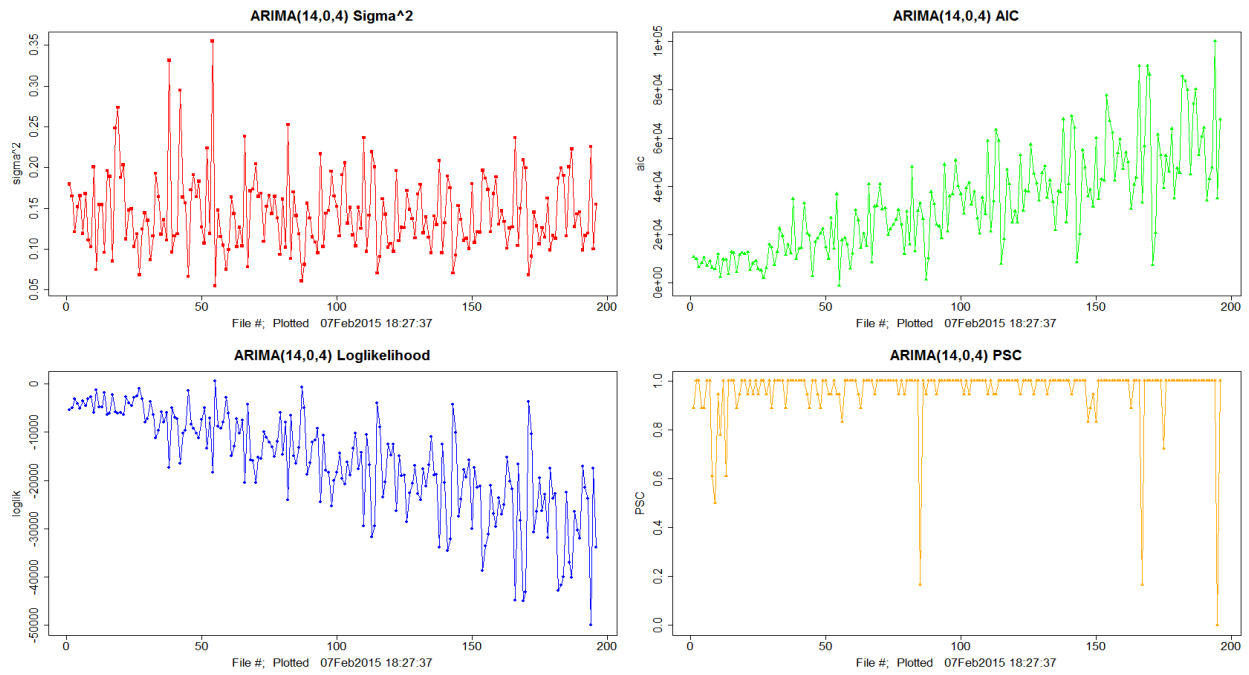


Figure 27: Example of the arima(14,0,4) performance (σ^2 , $\log - \text{likelihood}$, AIC and Percentage of Significant Coefficients (PSC)) over each data.

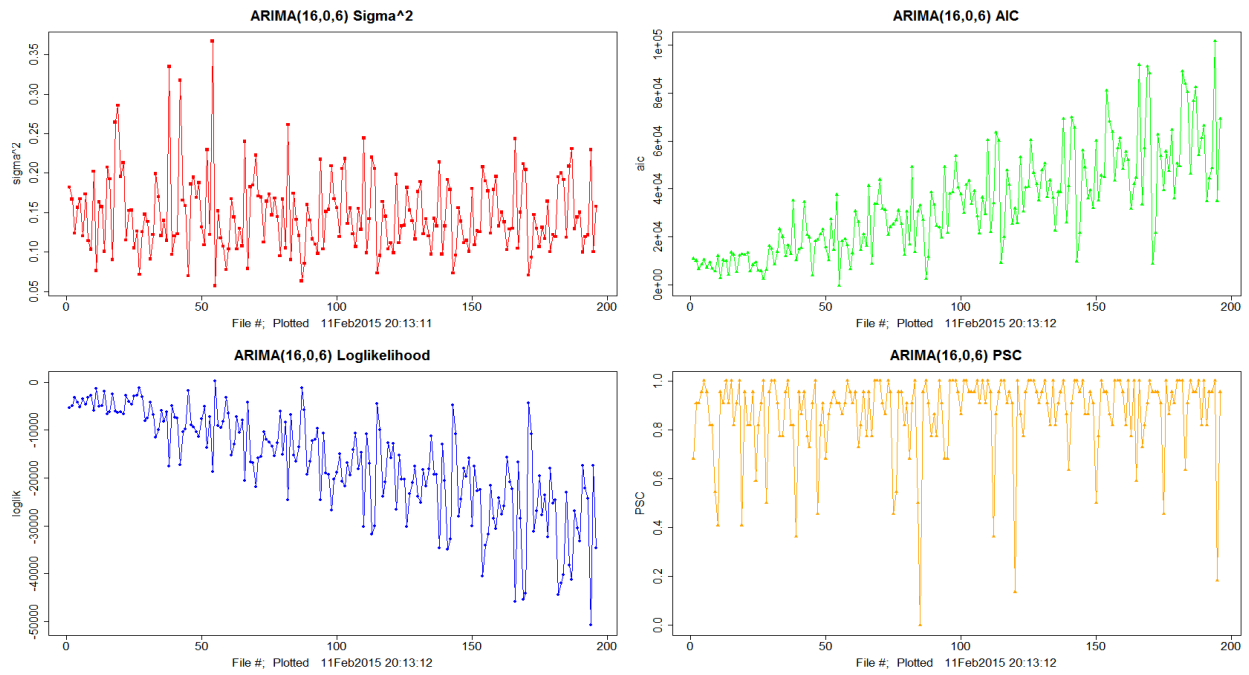


Figure 28: Example of the arima(16,0,4) performance (σ^2 , $\log - \text{likelihood}$, AIC and Percentage of Significant Coefficients (PSC)) over each data.

Model Adequacy Check for arima(20,0,6) Model

An examination of the arima(20,0,6) model adequacy was conducted by the following three steps. Similar examinations are carried out for other arima(p,0,q) models, and the results are very consistent.

First, check the arima(20,0,6) residuals with p-value vs lag plot. $\text{Max}(\text{lag}) = 35$ shown as illustrated. For 1-flight-data, all p-values are around 1 (0.92 ~1); for 4-combined-flights-data, lag (1, 25) have p-values almost equal to 1; for 7-combined-flight-data, lag (1, 23) have p-values almost equals to 1. According to *Box.test{stats}* help file, the p-value represents the likelihood to be against independent. Therefore, the p-value near 1.0 indicates that the residual is independent, not against independence. A p-value near 0 would indicates that the residual is not independent. This is different from the normal meaning of p-value for Linear Regression.

The R codes used are:

```
x.fit <- arima(x, order = c(p, 0, q), optim.method = "Nelder – Mead")  
p[i] <- Box.test(x.fit$residuals, i, type = 'Ljung – Box')$p.value.
```

Second, check the arima(20,0,6) residuals with Classical 4 Plots: residuals in order; ACF of residuals; histogram of residuals, and normal probability plot. The run-order of residuals does not show any particular pattern. The ACF of residuals illustrates only a first order is significant (this is very desirable). The Histogram plot shows that residuals are centered around zero and that all are distributed within the (-1,+1) region (this is desirable too). The Normal Probability Plot reveals to be within the (-1.5, +1.5) region, and the residuals are normally distributed. The ideal situation is to be normally distributed within a (-2,+2) region. This check demonstrates the goodness of the residuals.

The R codes used are:

```
plot (x.fit$residuals)
acf (x.fit$residuals)
hist(x.fit$residuals,breaks = 100,prob = T) | lines(density(x.fit$residuals))
qqnorm(x.fit@residuals) | qqline(x.fit@residuals).
```

Third, check the arima(20,0,6) residuals with the R built-in function *tsdiag{stats}*. It's a combination of Standardized Residuals, ACF of Residuals, and p-values of Ljung-Box Statistics. The run-order of standardized residuals does not show any particular pattern. That is, the ACF of residuals shows only first order is significant (this is good), and the p-values of Ljung-Box statistics showed near value of 1.

The R code used is:

```
tsdiag (x.fit$residuals).
```

The following are some example plots from 1-flight-data, 4-combined-flights-data, and 7-combined-flights-data.

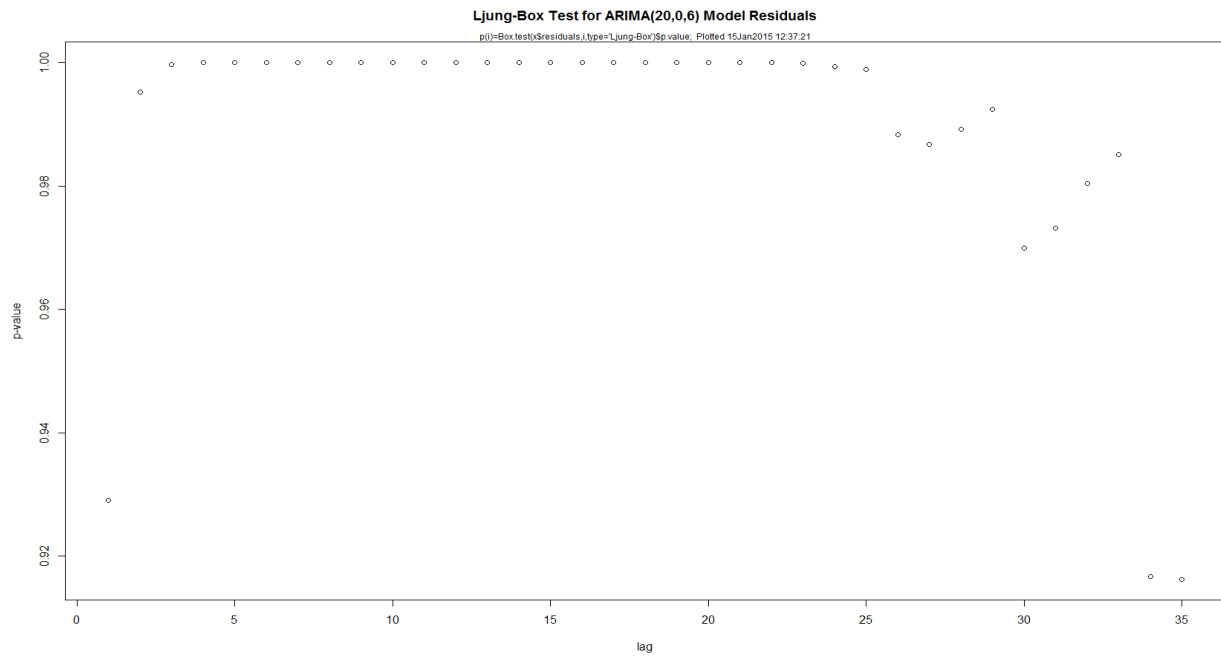


Figure 29: Model Adequacy check - one flight. The ARIMA(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).

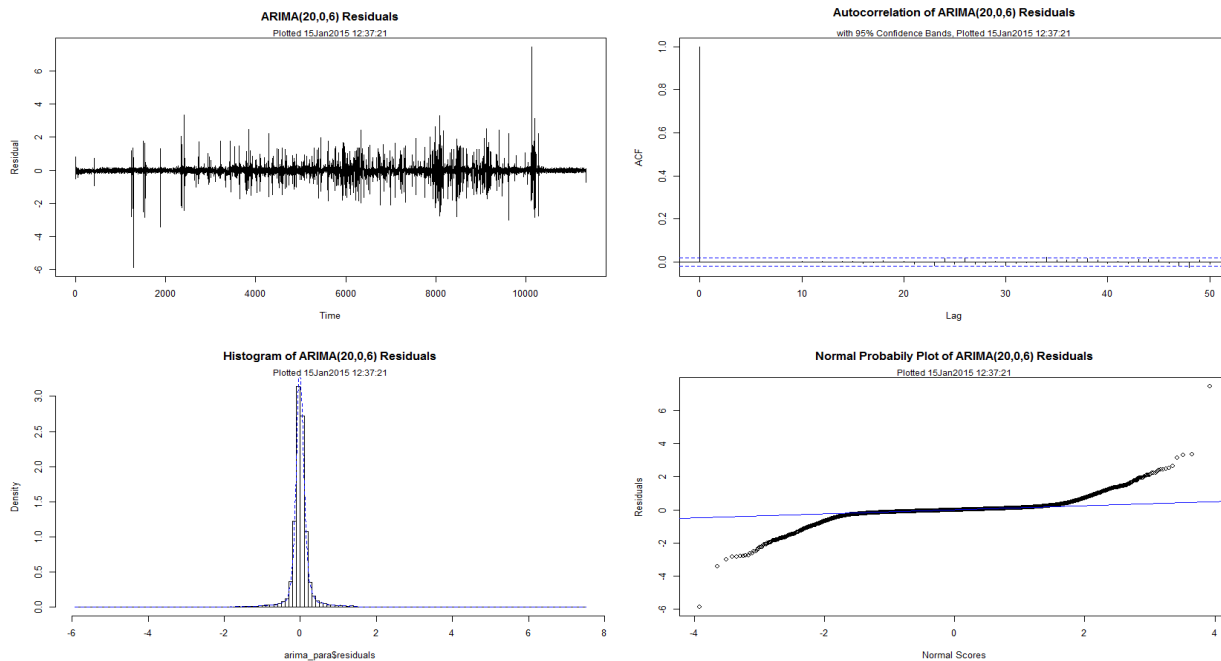


Figure 30: Model Adequacy check - one flight. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.

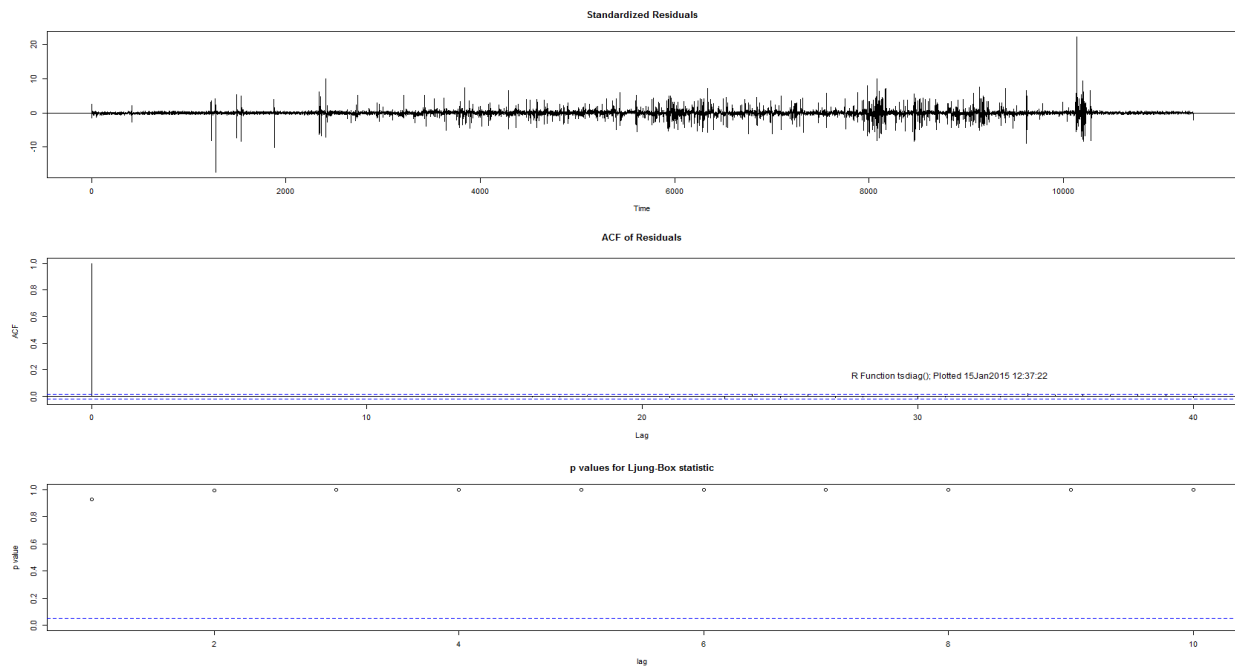


Figure 31: Model Adequacy check - one flight. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().

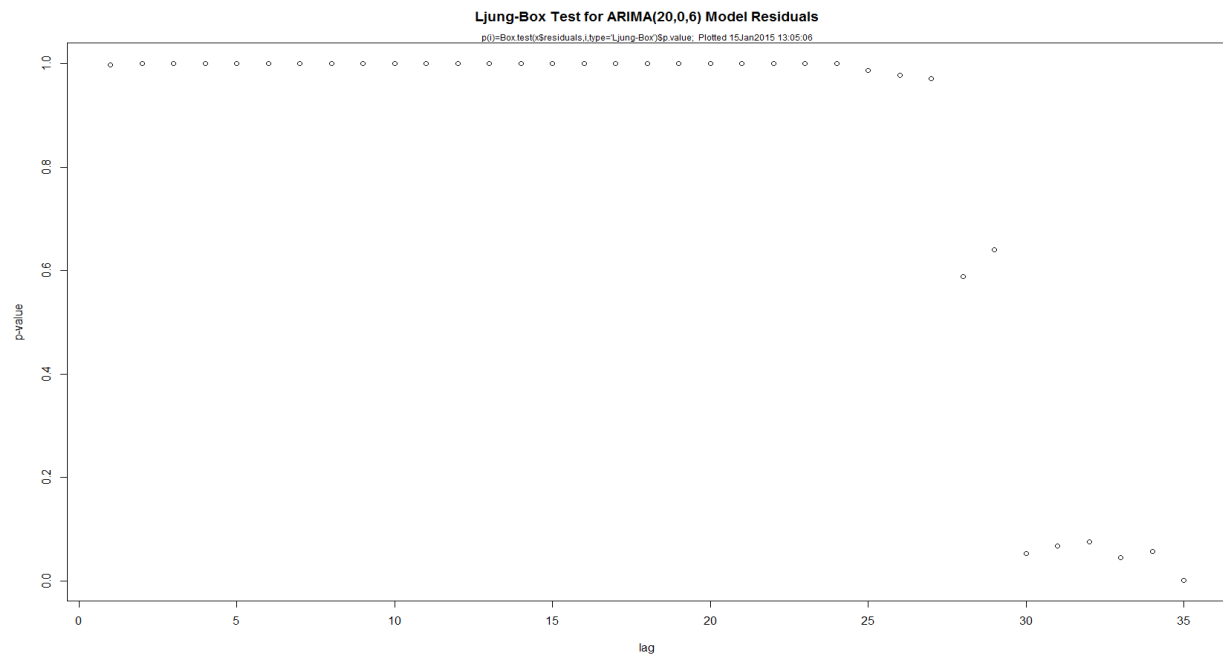


Figure 32: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).

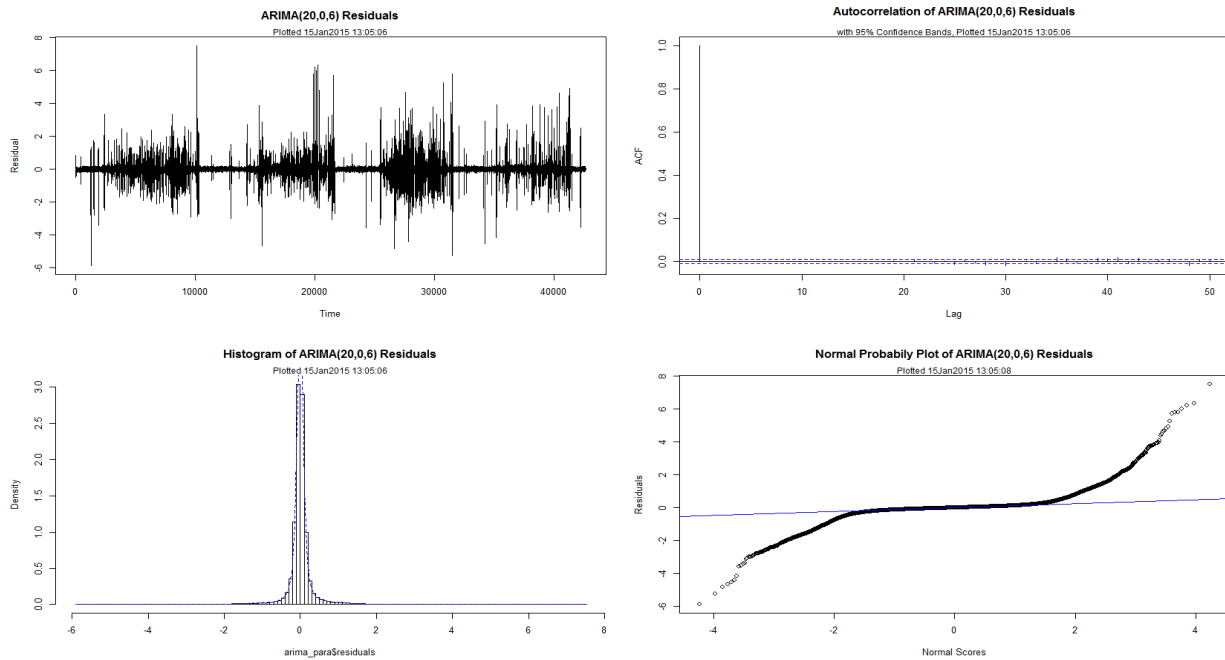


Figure 33: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.

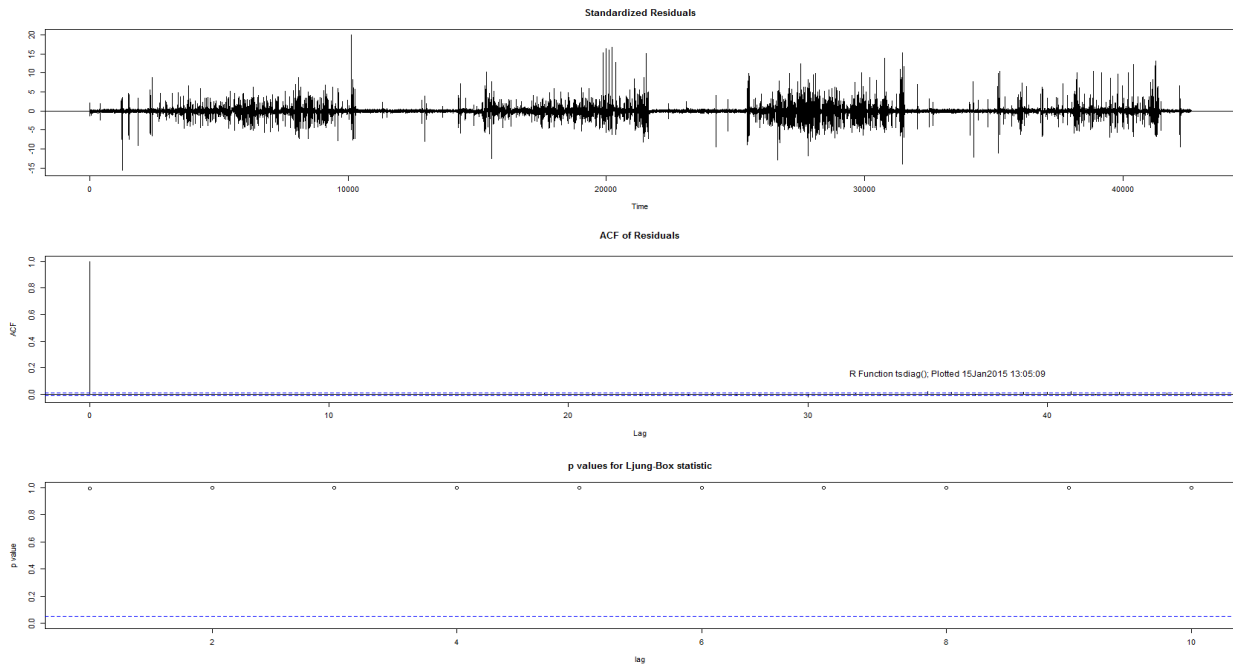


Figure 34: Model Adequacy check - four flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().

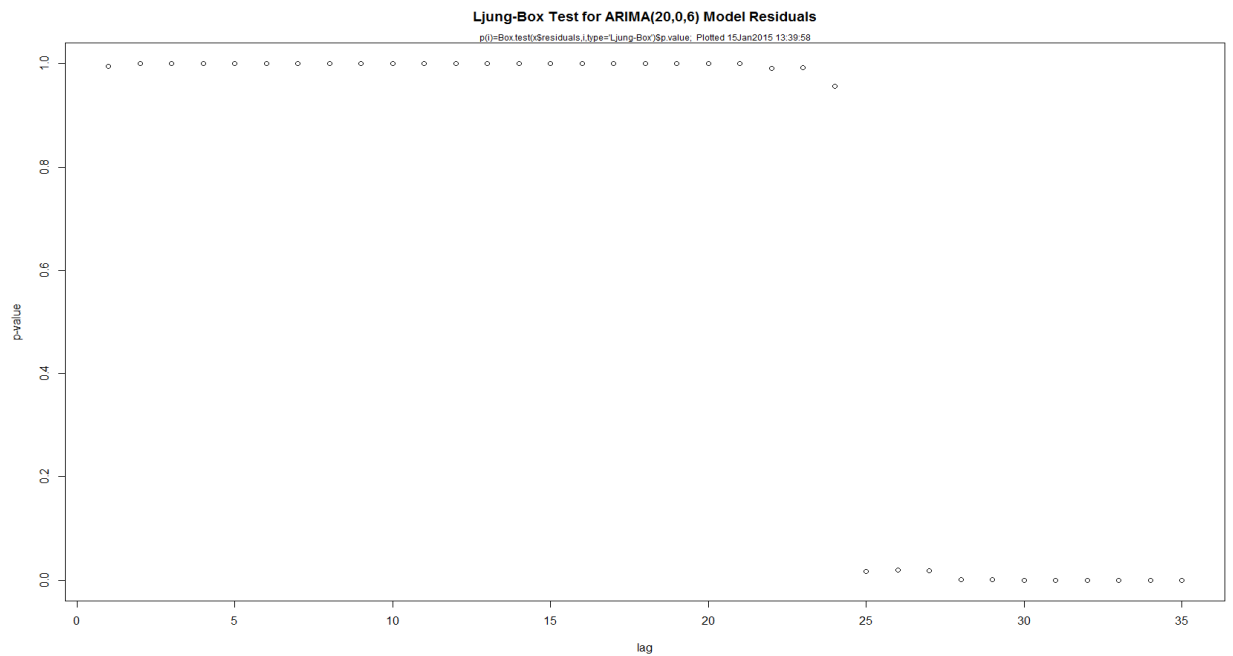


Figure 35: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, and p-values of Ljung-Box Test versus lag (max of 35).

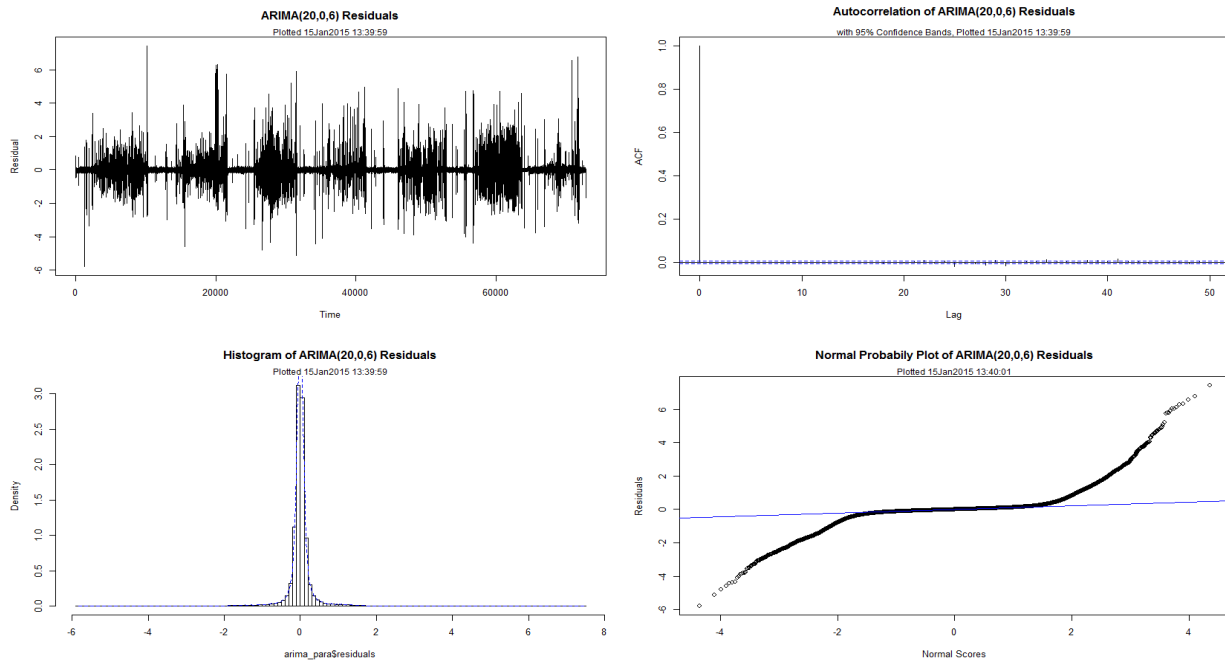


Figure 36: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, in Classical 4-plots.

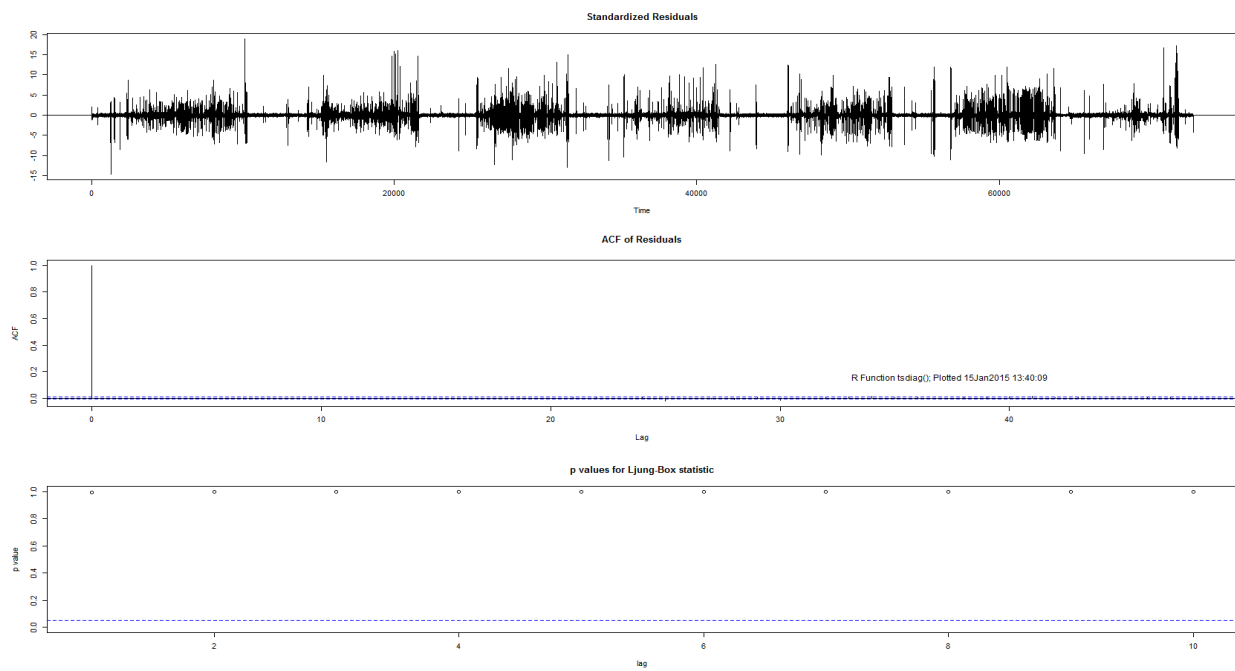


Figure 37: Model Adequacy check - seven flights. The arima(20,0,6) for first order differenced ts Variable Geometry(VG) data, using R built-in function tsdiag().

Model Adequacy Check for Probability Model of Compressor Stall Fault Event

The R built-in plotting function $plot(lm\{stats\})$ have been used to check the adequacy of the fitted Linear Regression Model. Furthermore, R^2 and R_{adj}^2 have been retrieved from the results of R built-in function $lm\{stats\}$. Additionally the Model's p-values (not the p-values of the fitted model's coefficients) were calculated by using following R codes:

```
fit.lm <- lm(Fault ~ ., data = DATA, na.action = NULL)
```

```
 $R^2$  <- summary(fit.lm)$r.squared
```

```
 $R_{adj}^2$  <- summary(fit.lm)$adj.r.squared
```

```
Standard Error <- summary(fit.lm)$sigma
```

```
Model_p - value <- pf(x[1], x[2], x[3], lower.tail = FALSE)
```

Where $x <- summary(fit.lm)$fstatistic$; $pf\{stats\}$ is the F probability distribution function built-in R programming.

The following plots show the adequacy check for the “best fit” based upon the current selection of sample data. The LRM result would be better if the sample size were increased. According to

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

where n is the sampling size, p is the number of regressor ($= p + q$ in $arima(p,0,q)$), a larger sample size would be very helpful to obtain a larger R_{adj}^2 if the same R^2 can be somewhat retained.

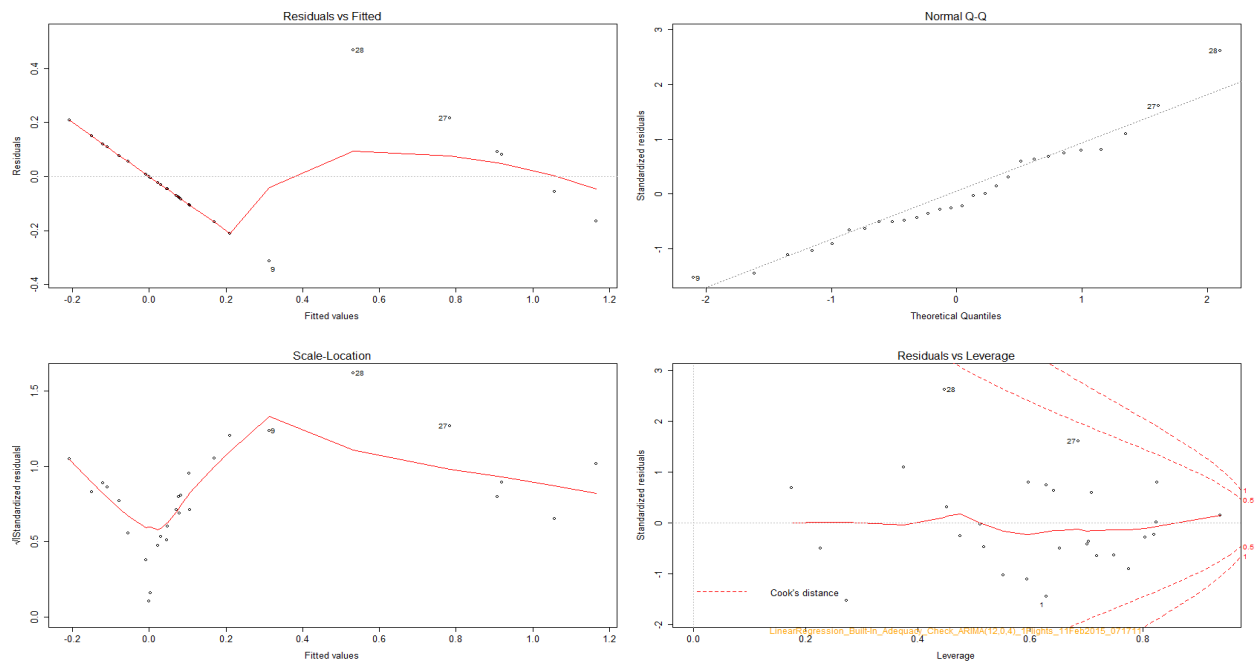


Figure 38: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for one-flight-data with `arima(12,0,4)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,2)	0.3528	-0.3443	0.4845	0.8902
arima(12,4)	0.8656	0.67	0.24	0.008221
arima(12,6)	0.6282	-0.1154	0.4413	0.6383
arima(14,2)	0.7846	0.4712	0.3039	0.06376
arima(14,4)	0.6257	-0.1229	0.4428	0.6452
arima(14,6)	0.5488	-0.7405	0.5513	0.9366
arima(16,2)	0.8493	0.548	0.2809	0.05775
arima(16,4)	0.7783	0.145	0.3864	0.4134
arima(16,6)	0.6217	-1.043	0.5972	0.951

Figure 39: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the one-flight-data, the “best fit” is `arima(12,0,4)`.

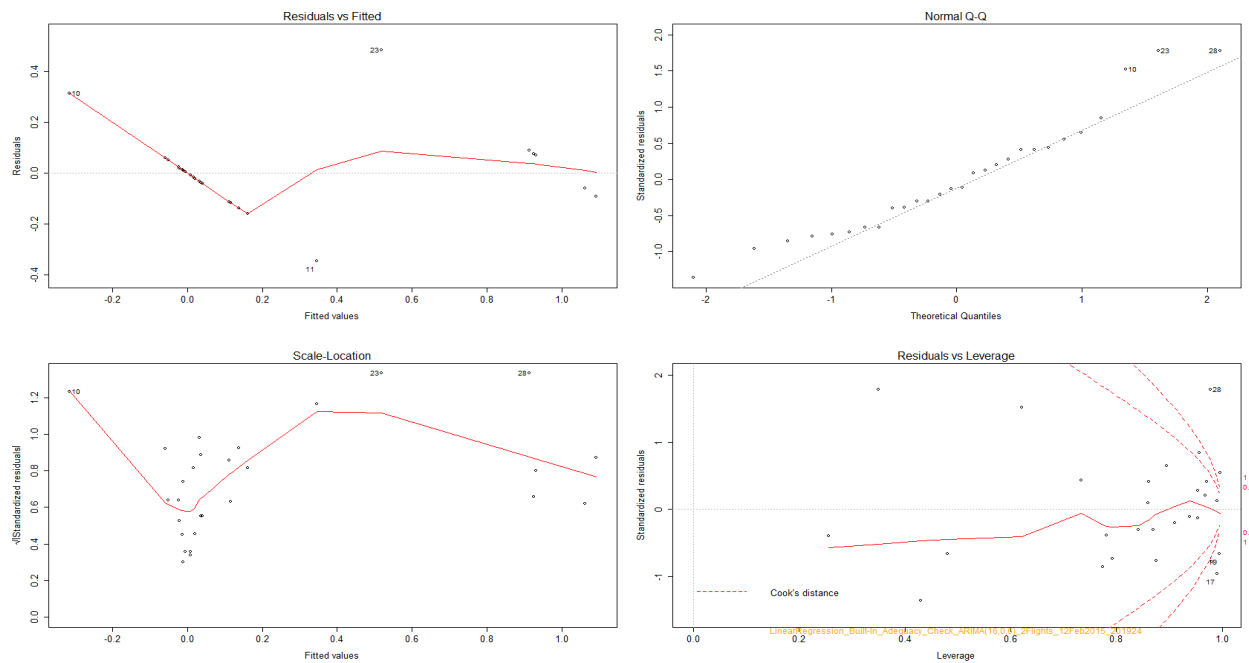


Figure 40: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for two-combined-flight-data with `arima(16,0,6)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
ARIMA(12,0,2)	0.4757	-0.089	0.4361	0.624
ARIMA(12,0,4)	0.4276	-0.405	0.4953	0.8905
ARIMA(12,0,6)	0.757	0.271	0.3568	0.2522
ARIMA(14,0,2)	0.6573	0.1587	0.3833	0.3262
ARIMA(14,0,4)	0.5764	-0.2707	0.471	0.7676
ARIMA(14,0,6)	0.7919	0.1972	0.3744	0.3669
ARIMA(16,0,2)	0.4798	-0.5606	0.522	0.9224
ARIMA(16,0,4)	0.632	-0.4196	0.4979	0.8248
ARIMA(16,0,6)	0.88	0.3521	0.3363	0.2995

Figure 41: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the two-combined-flight-data, the “best fit” is `arima(16,0,6)`.

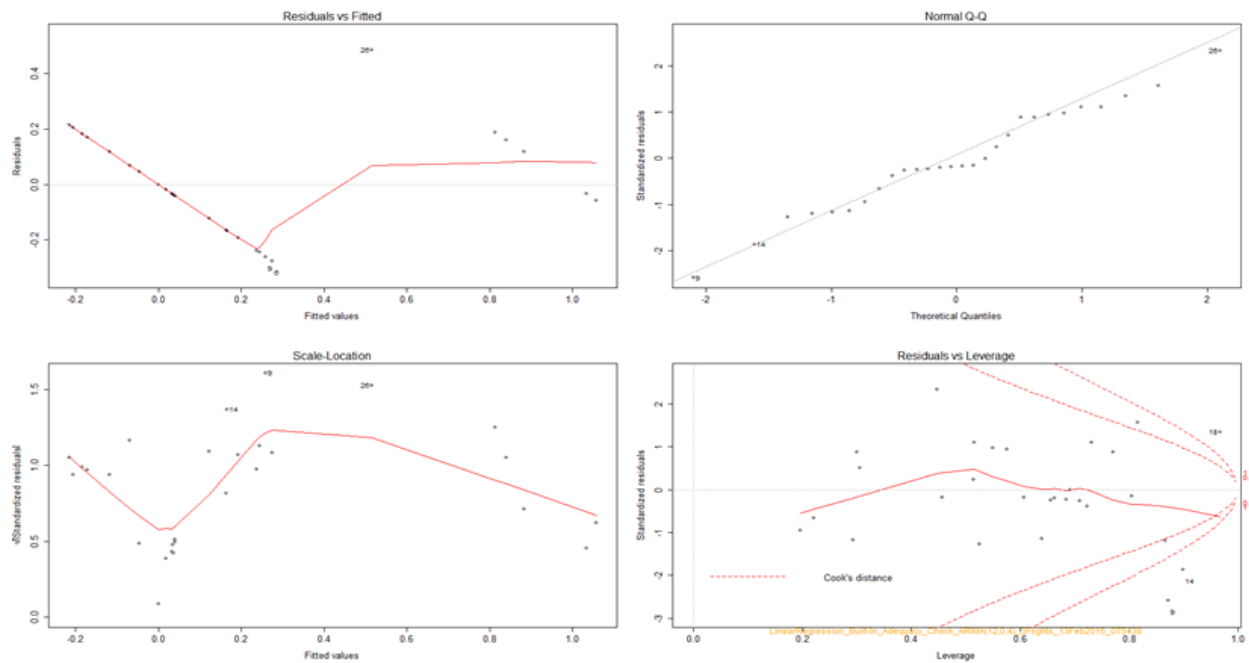


Figure 42: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for three-combined-flight-data with `arima(12,0,4)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,0,2)	0.5135	-0.01048	0.42	0.5172
arima(12,0,4)	0.8179	0.5531	0.2793	0.03177
arima(12,0,6)	0.7241	0.1724	0.3801	0.3484
arima(14,0,2)	0.7329	0.3444	0.3383	0.1445
arima(14,0,4)	0.7888	0.3664	0.3326	0.1699
arima(14,0,6)	0.8827	0.5476	0.2811	0.09659
arima(16,0,2)	0.638	-0.08585	0.4354	0.6105
arima(16,0,4)	0.7958	0.2124	0.3708	0.3533
arima(16,0,6)	0.6378	-0.9558	0.5844	0.9384

Figure 43: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the three-combined-flight-data, the “best fit” is `arima(12,0,4)`.

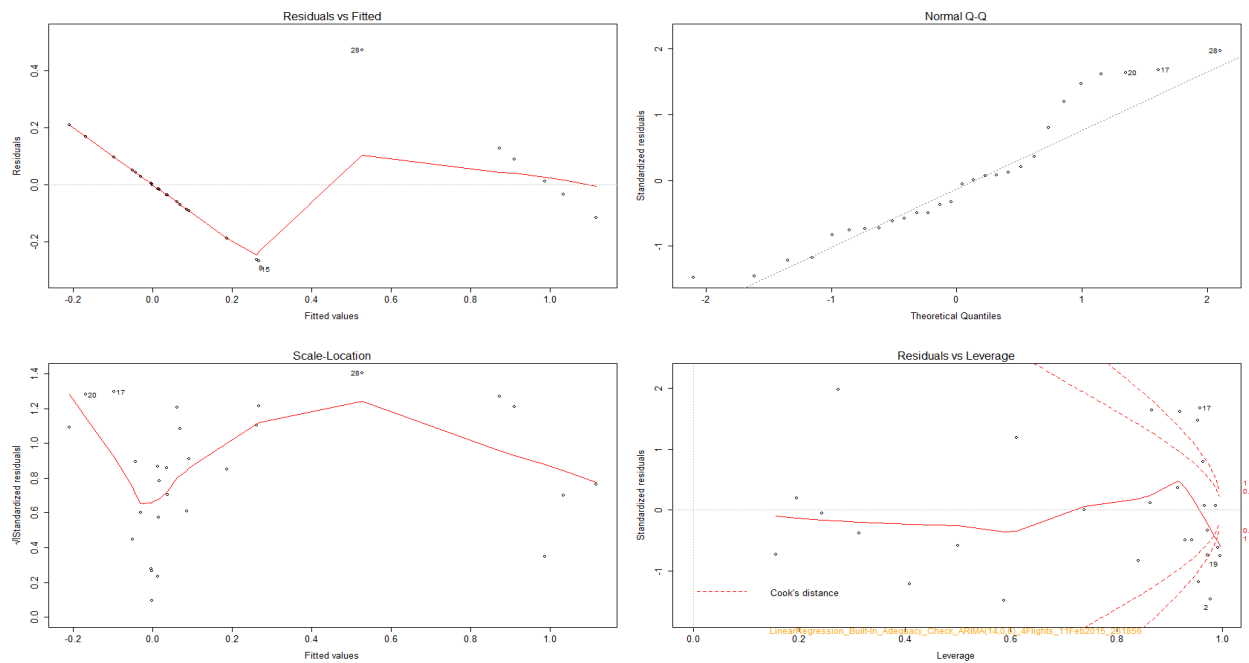


Figure 44: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for four-combined-flight-data with `arma(14,0,6)`.

arma_order	r_squared	adj_r_squared	standard_error	model_p-value
arma(12,2)	0.6894	0.3549	0.3356	0.1007
arma(12,4)	0.3939	-0.4876	0.5096	0.9303
arma(12,6)	0.4542	-0.6375	0.5347	0.9459
arma(14,2)	0.4902	-0.2513	0.4674	0.7807
arma(14,4)	0.6323	-0.1031	0.4389	0.6268
arma(14,6)	0.8826	0.5473	0.2812	0.09674
arma(16,2)	0.3989	-0.8034	0.5611	0.9778
arma(16,4)	0.5448	-0.7556	0.5537	0.9399
arma(16,6)	0.69	-0.6738	0.5406	0.8778

Figure 45: The examination of Linear Regression Model (LRM) adequacy based on various `arma(p,0,q)` are shown here. For the four-combined-flight-data, the “best fit” is `arma(14,0,6)`.

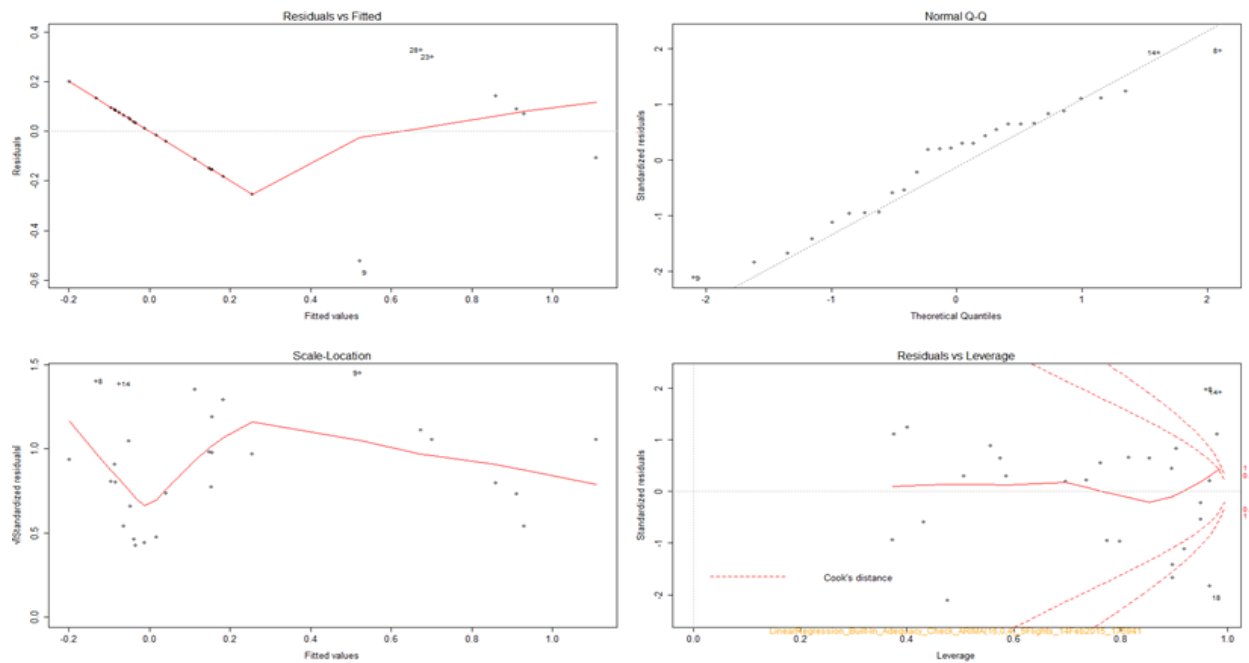


Figure 46: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for five-combined-flight-data with `arima(16,0,4)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,0,2)	0.5434	0.05175	0.4069	0.4312
arima(12,0,4)	0.5486	-0.108	0.4398	0.6385
arima(12,0,6)	0.4357	-0.6928	0.5437	0.959
arima(14,0,2)	0.3695	-0.5475	0.5198	0.9519
arima(14,0,4)	0.5583	-0.325	0.481	0.8054
arima(14,0,6)	0.7863	0.1756	0.3794	0.386
arima(16,0,2)	0.5525	-0.3426	0.4842	0.8168
arima(16,0,4)	0.8261	0.3291	0.3422	0.2525
arima(16,0,6)	0.7601	-0.2952	0.4756	0.733

Figure 47: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the five-combined-flight-data, the “best fit” is `arima(16,0,4)`.

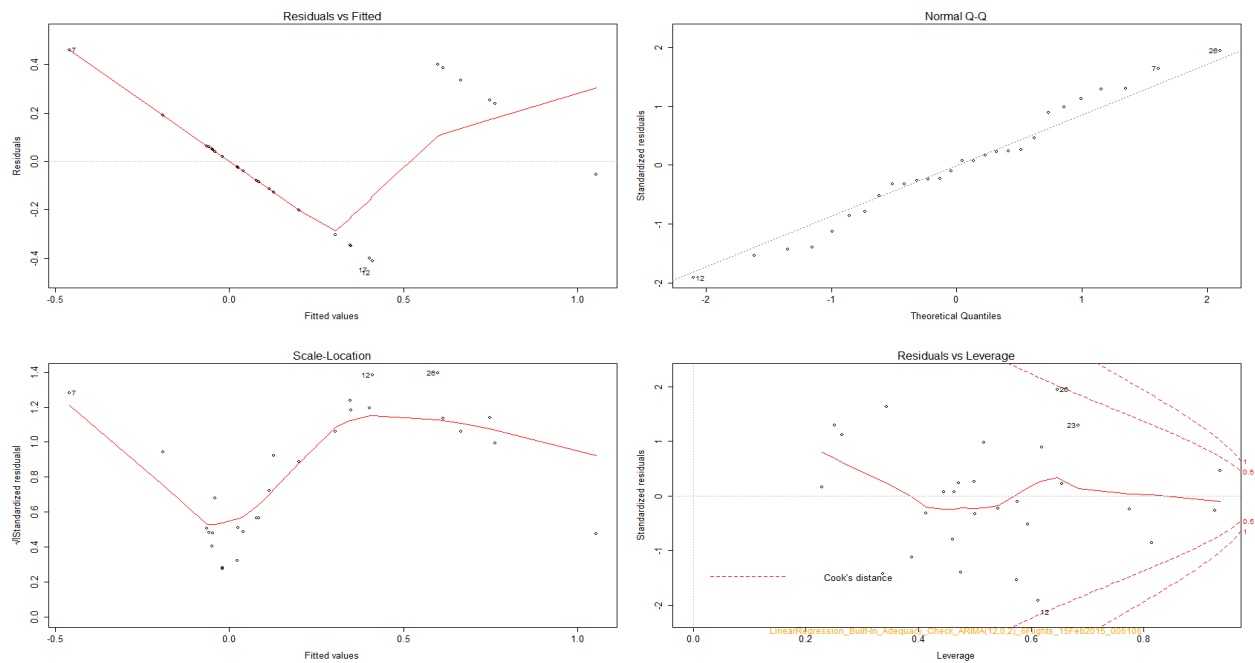


Figure 48: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for six-combined-flight-data with `arima(12,0,2)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,0,2)	0.6694	0.3134	0.3462	0.1318
arima(12,0,4)	0.5419	-0.1245	0.4431	0.6565
arima(12,0,6)	0.6664	-0.0008468	0.418	0.5267
arima(14,0,2)	0.5407	-0.1273	0.4437	0.6595
arima(14,0,4)	0.4856	-0.5433	0.5191	0.9163
arima(14,0,6)	0.7272	-0.05217	0.4286	0.5834
arima(16,0,2)	0.6892	0.06768	0.4035	0.4563
arima(16,0,4)	0.6659	-0.2886	0.4743	0.753
arima(16,0,6)	0.7802	-0.1871	0.4553	0.6756

Figure 49: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the six-combined-flight-data, the “best fit” is `arima(12,0,2)`.

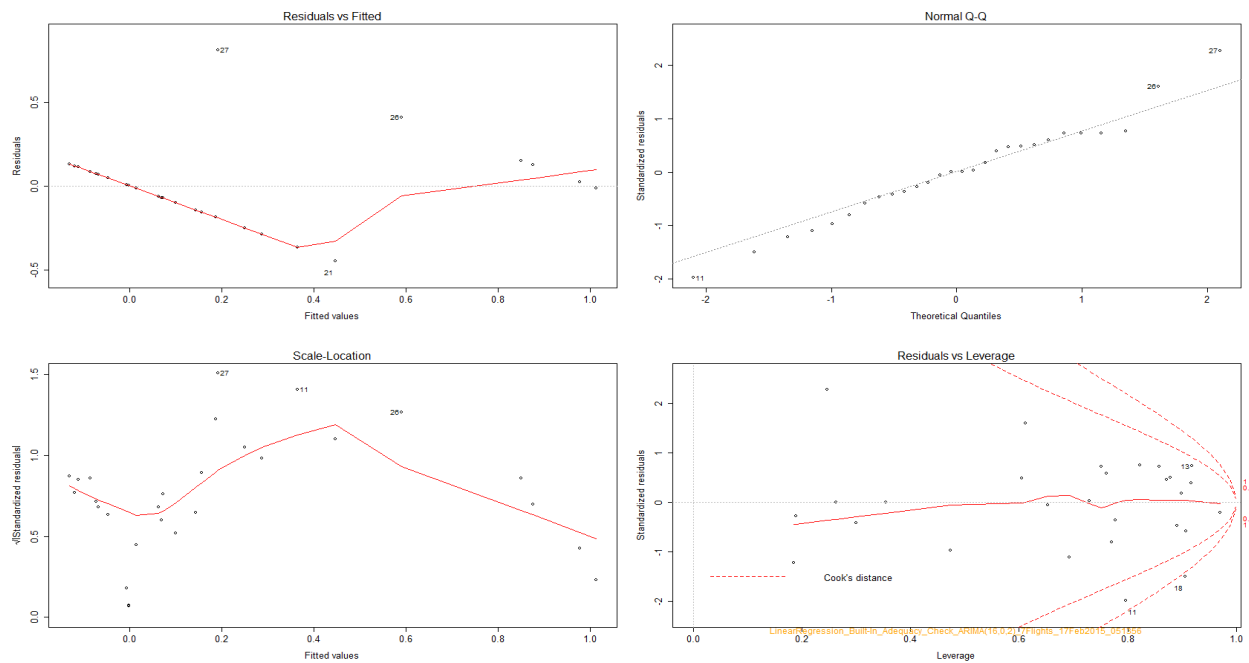


Figure 50: Check Linear Regression Model (LRM) adequacy with R built-in function `plot(lm{stats})`. This is an example for seven-combined-flight-data with `arima(16,0,2)`.

arima_order	r_squared	adj_r_squared	standard_error	model_p-value
arima(12,0,2)	0.4666	-0.1077	0.4398	0.6485
arima(12,0,4)	0.4723	-0.2952	0.4755	0.8169
arima(12,0,6)	0.6658	-0.002485	0.4184	0.5284
arima(14,0,2)	0.2671	-0.799	0.5605	0.9937
arima(14,0,4)	0.5095	-0.4715	0.5069	0.8868
arima(14,0,6)	0.6191	-0.4694	0.5065	0.8479
arima(16,0,2)	0.6808	0.04247	0.4089	0.4824
arima(16,0,4)	0.4908	-0.9641	0.5856	0.9735
arima(16,0,6)	0.676	-0.7495	0.5527	0.8974

Figure 51: The examination of Linear Regression Model (LRM) adequacy based on various `arima(p,0,q)` are shown here. For the seven-combined-flight-data, the “best fit” is `arima(16,0,2)`.

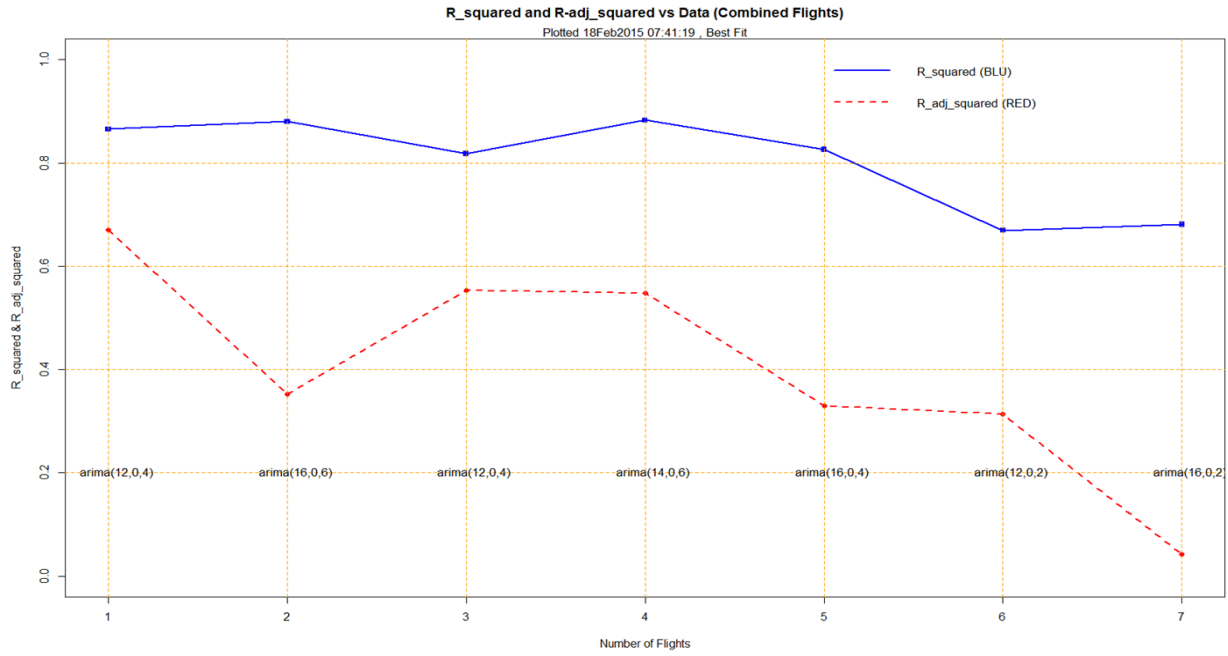


Figure 52: Engine Stall Probability Model Adequacy Check – R^2 and R_{adj}^2 of “Best Fit” Models versus the number of combined flights.

These are the “best fit” LRM observed, based on current flight data samples, and utilizing

$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$, where n is sampling size, p is number of regressor (= $p + q$ in

arima($p,0,q$)). Once again, a larger sample size would be very helpful to obtain a larger R_{adj}^2 if

the same R^2 can be somewhat retained.

Relevant Research – Application of Weight Filter for Times Series Data

It will be beneficial to use a Weight Filter in order to achieve a more concise PACF and/or ACF, and thereby lower the order of $\text{arma}(p,0,q)$ to assist in shortening the calculation run times.

For example, currently using a normal computer (CPU - AMD Athlon64/3400/2.2GHz Memory - 2.0GB), running the R script “Linear_Regression_Model_FOR_EngineStall_Scan.R” takes 32:58:24 for the 7-combined-flights-data using the 22 fault-Negative samples plus 6 fault-Positive samples. Compared to the 6-combined-flights-data, processing those 28 samples would take 28:21:56. Finally for the 3-combined-flights-data, processing those 28 samples would only take 13:56:04.

Three types of Weight filters were used in an attempt to sharpen the time series data in an effort to achieve the stated objective. The 3 filter types were experimented in arma model with 4 combined-flights-data of VG.

Unfortunately, the current format of filters used did not yield any sign of reaching the desired objectives. In the Future Research and Discussion section of this study, another format of filter, the Kalman filter, is proposed to be employed.

For a centered Convolution filter which was used, the common mathematical expression is given by:

$$x.filter[i] = f[1] * x[i + o] + f[2] * x[i + (o - 1)] + \dots + f[(p - 1)] * x[i + o - p + 2] + f[p] * x[i + o - (p - 1)].$$

Where the original time series data is $x[i]$, and the filtered time series data is $x.filter[i]$, o is the offset $o=(p-1)/2$, and p is the number of filter points and must be an odd number under the

convolution filter type.

The filter itself is represented by:

$$f[i] = f[1] * x[i + o] + f[2] * x[i + (o - 1)] + \dots + f[(p - 1)] * x[i + o - (p - 2)] + f[p] * x[i + o - (p - 1)]$$

In R, it is implemented as:

$$x.filter <- filter(x, c(f[1], f[2], \dots, f[(p - 1)], f[p], sides = 2))$$

Where the filter sharp $f[i]$ is formatted as

$$\sum_{i=1}^p f[i] = 1$$

In this way, the filtered data will meet two critical requirements. First it will be a time series, and second it will be a stationary set of data which is desired for an ARIMA model.

The results of the applied filter exercises are provided below.

Filter 15

For a 15 point filter, the following shape is experimented:

$$f[i] = \left[\frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{2}{27}, \frac{2}{27}, \frac{2}{27}, \frac{3}{27}, \frac{3}{27}, \frac{3}{27}, \frac{2}{27}, \frac{2}{27}, \frac{2}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27} \right]$$

The mathematical expression of the filtered time series data is:

$$x.filter[i] = f[1] * x[i + 7] + f[2] * x[i + 6] + \dots + f[8] * x[i] + \dots + f[14] * x[i - 6] + f[15] * x[i - 7]$$

The three types of filter results are shown in figure 23.

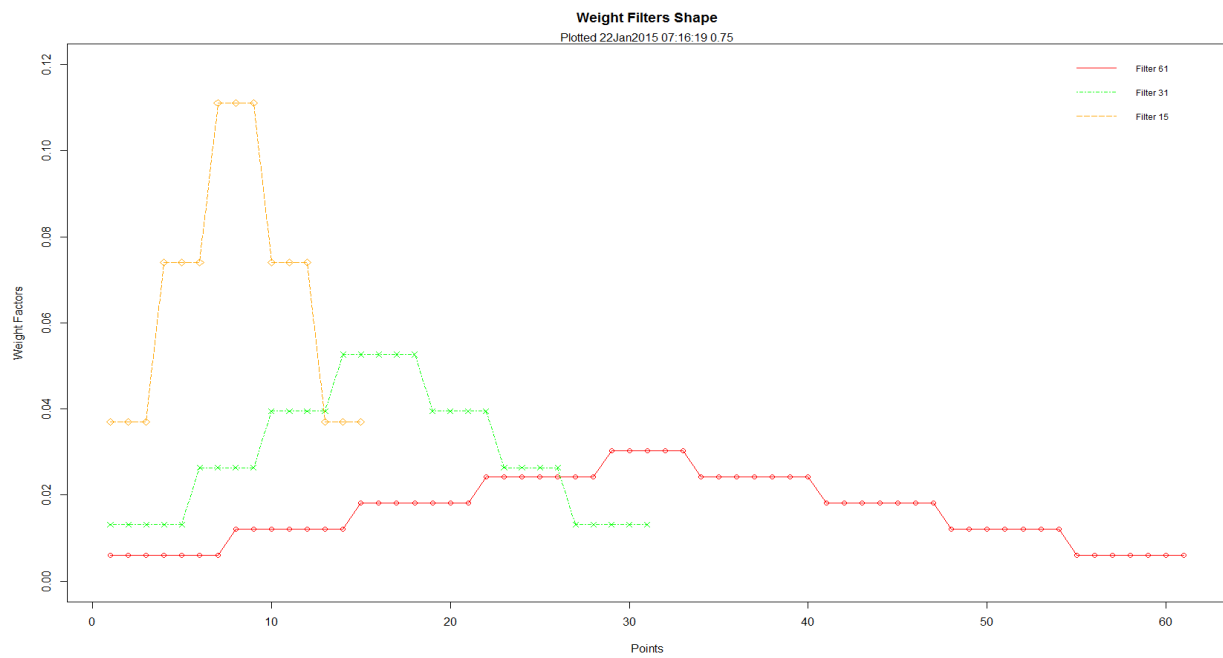


Figure 53: Shapes of the three filters experiments.

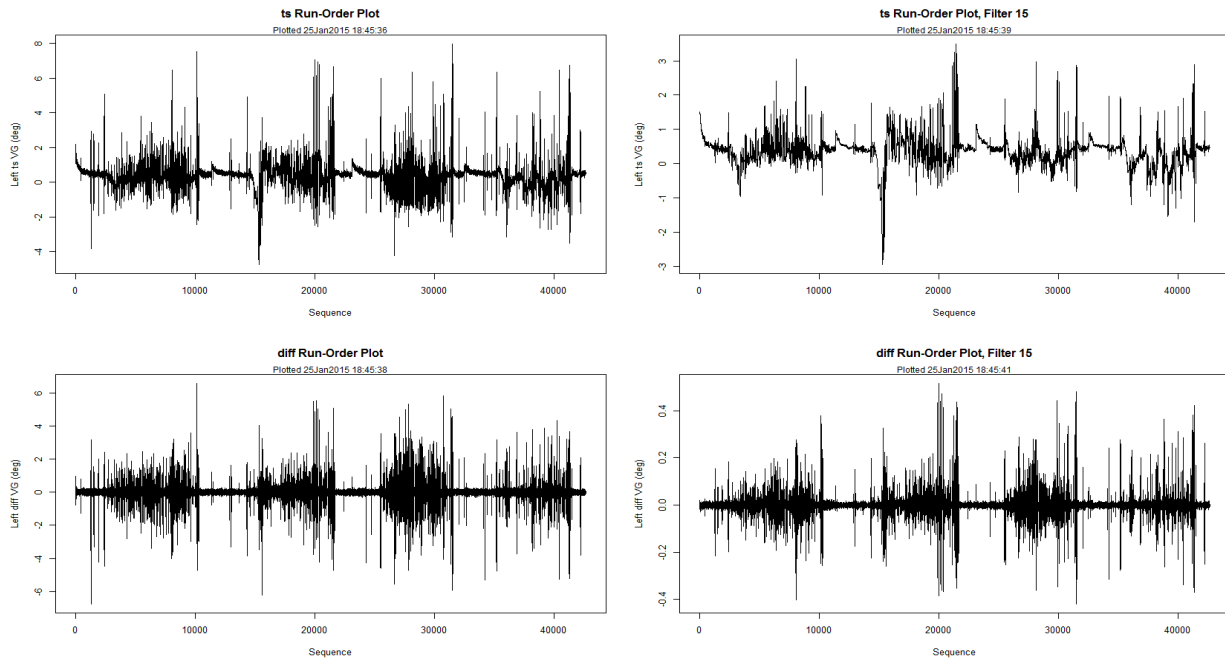


Figure 54: The effect of Filter 15 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.

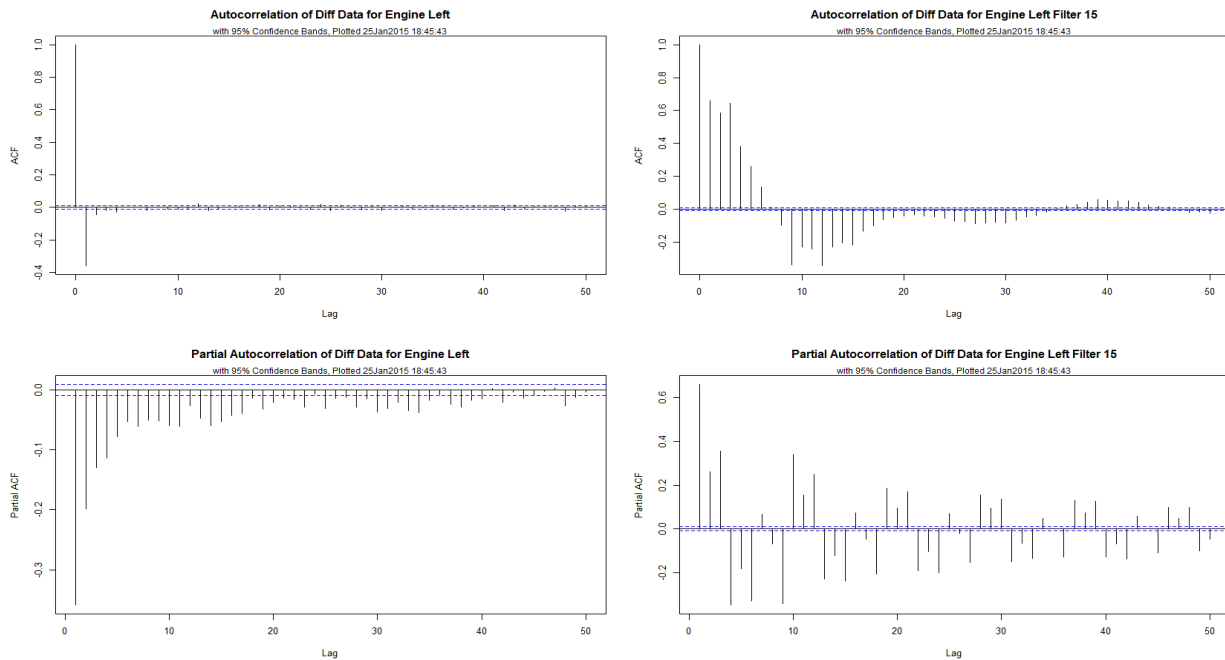


Figure 55: The effect of Filter 15 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.

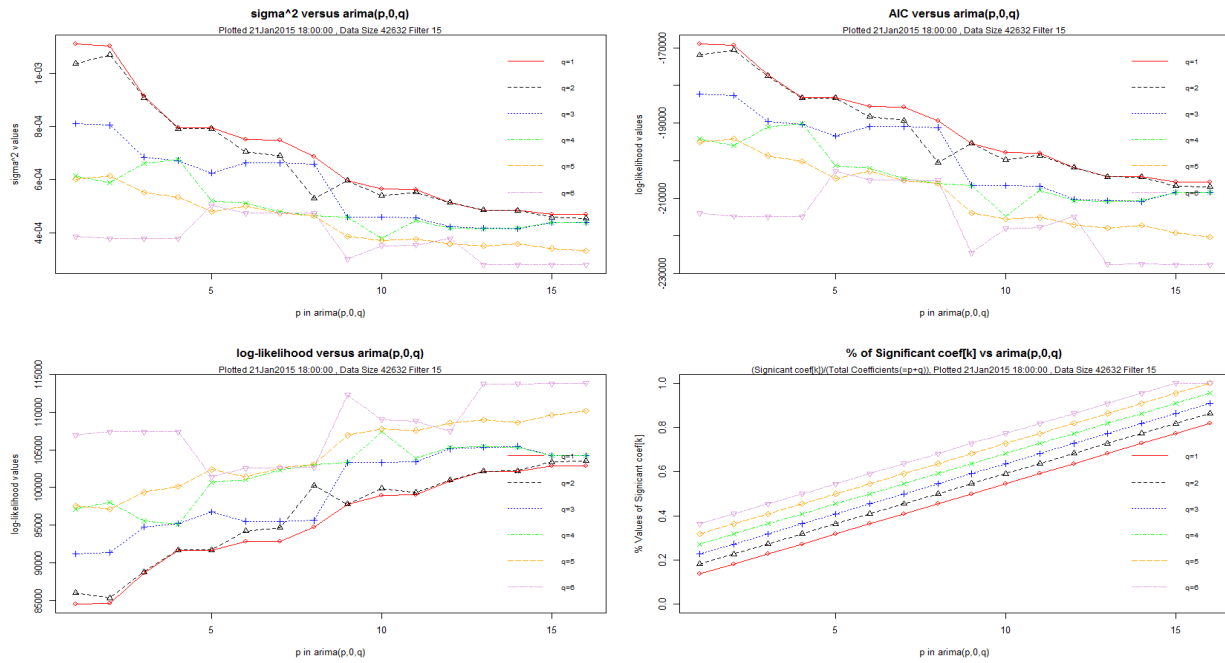


Figure 56: The effect of Filter 15. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of first order differenced Variable Geometry (VG) data.

Filter 31

For a 31 point filter, the following shape is experimented:

$$f[i] = \begin{bmatrix} \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \\ \frac{4}{76}, \frac{4}{76}, \frac{4}{76}, \frac{4}{76}, \frac{4}{76}, \\ \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \\ \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \frac{3}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{2}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76}, \frac{1}{76} \end{bmatrix}$$

Where $i = 1, 2, \dots, 30, 31$.

The mathematical expression of the filtered time series data is:

$$\begin{aligned} x.filter[i] = & f[1] * x[i + 15] + f[2] * x[i + 14] + \dots + f[16] * x[i] + \dots + f[30] * x[i - 14] \\ & + f[31] * x[i - 15] \end{aligned}$$

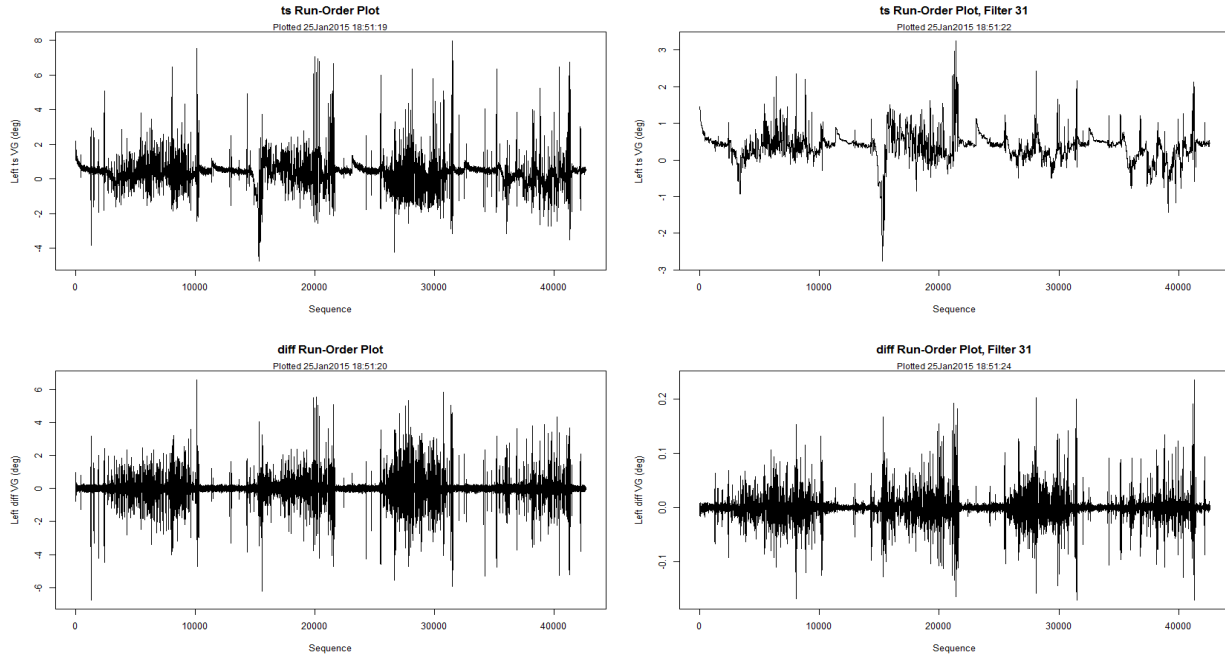


Figure 57: The effect of Filter 31 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.

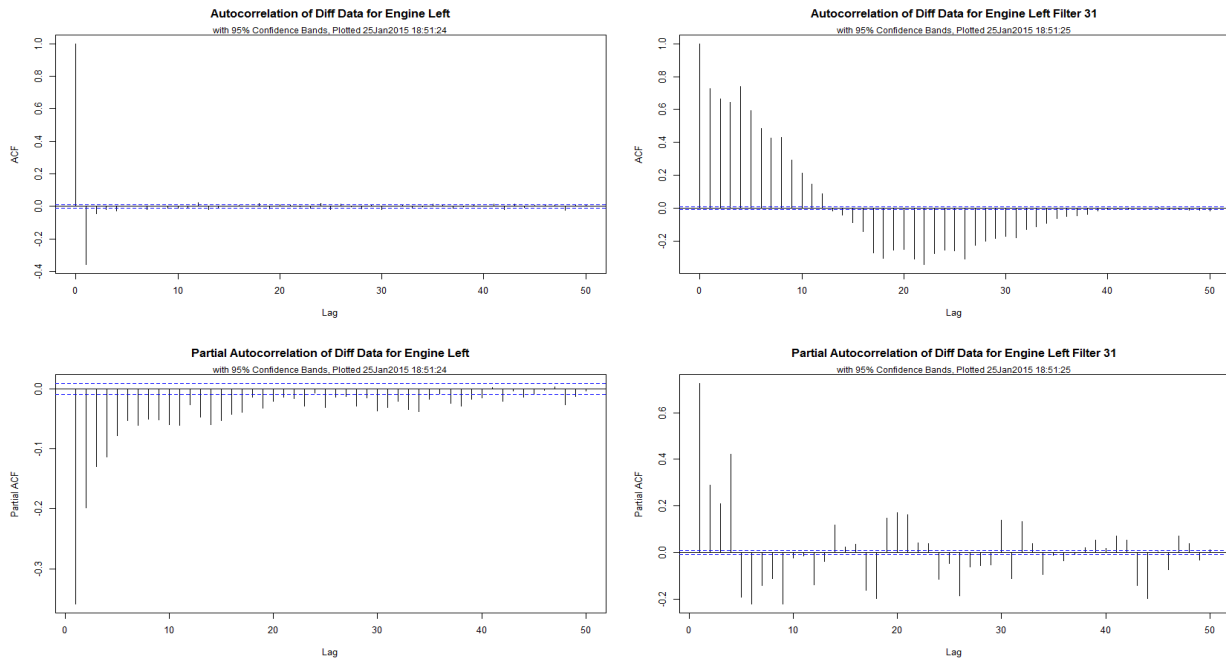


Figure 58: The effect of Filter 31 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.

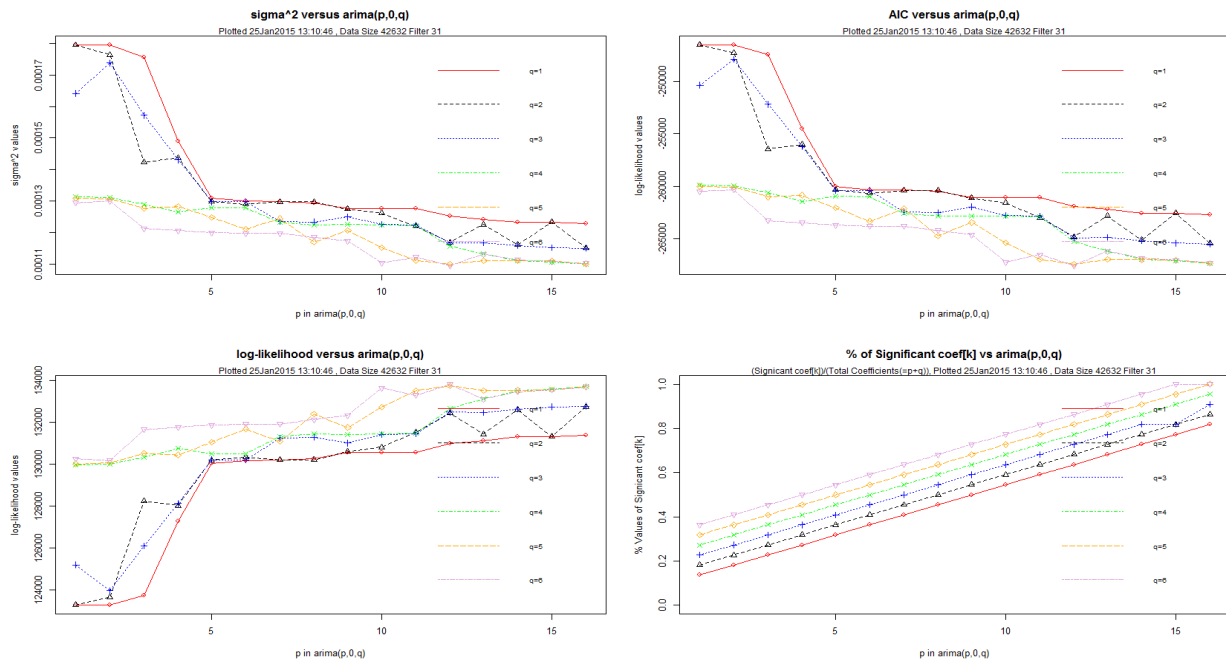


Figure 59: The effect of Filter 31. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of the first order differenced Variable Geometry (VG) data.

Filter 61

For a 61 point filter, the following shape is experimented:

$$f[i] = \begin{bmatrix} \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \\ \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \\ \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \\ \frac{5}{165}, \frac{5}{165}, \frac{5}{165}, \frac{5}{165}, \frac{5}{165}, \\ \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{4}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \frac{3}{165}, \\ \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \\ \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{2}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \frac{1}{165}, \\ \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165}, \frac{165}{165} \end{bmatrix}$$

Where $i = 1, 2, \dots, 60, 61$.

The mathematical expression of the filtered time series data is:

$$\begin{aligned} x.filter[i] &= f[1] * x[i + 30] + f[2] * x[i + 29] + \dots + f[31] * x[i] + \dots + f[60] * x[i - 29] \\ &\quad + f[61] * x[i - 30] \end{aligned}$$

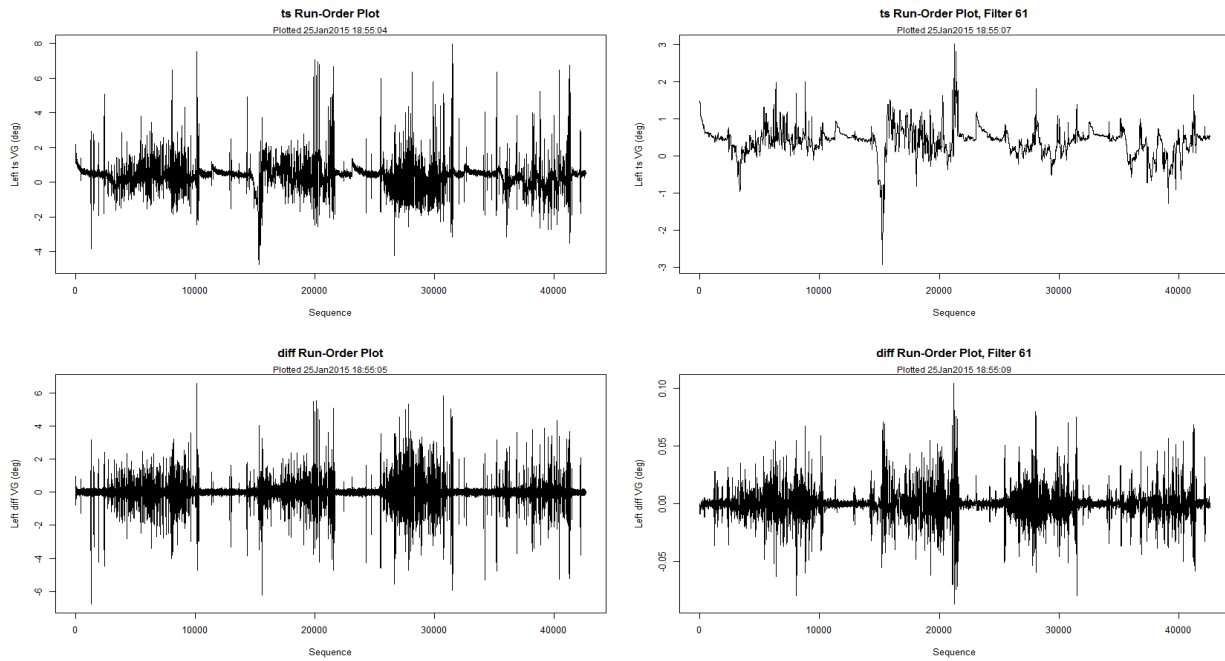


Figure 60: The effect of Filter 61 to the original Variable Geometry (VG) data and the first order differenced VG data which includes four flights.

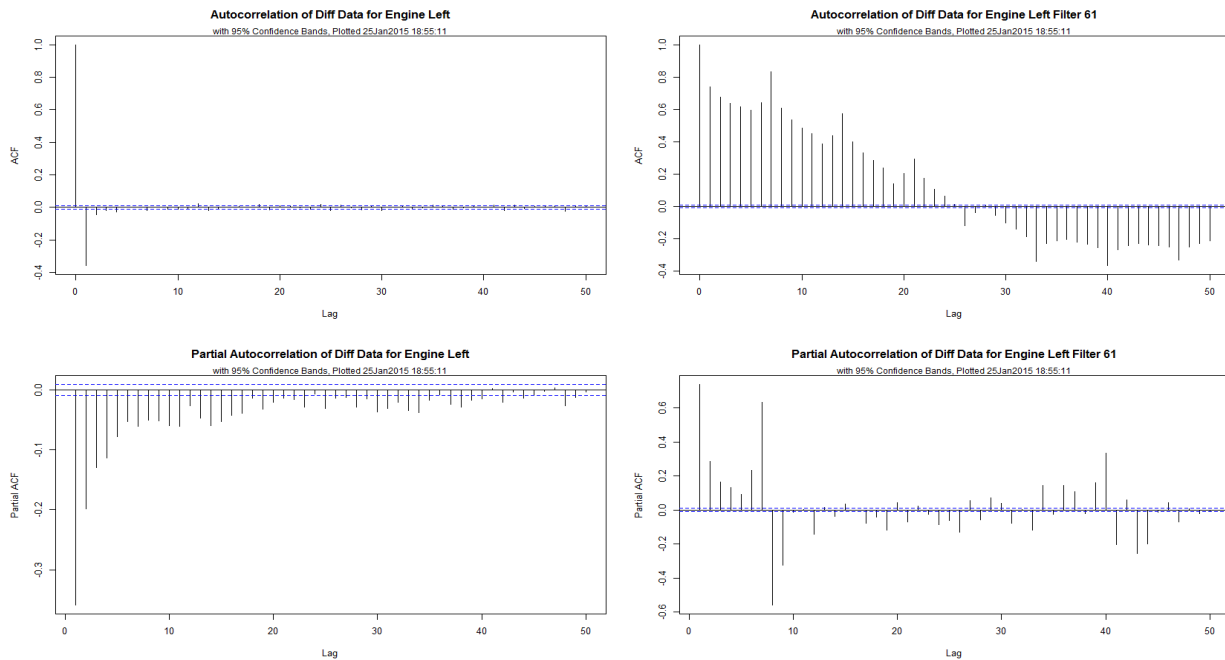


Figure 61: The effect of Filter 61 to the Autocorrelation Function and Partial Autocorrelation Function of the first order differenced Variable Geometry (VG) data which includes four flights.

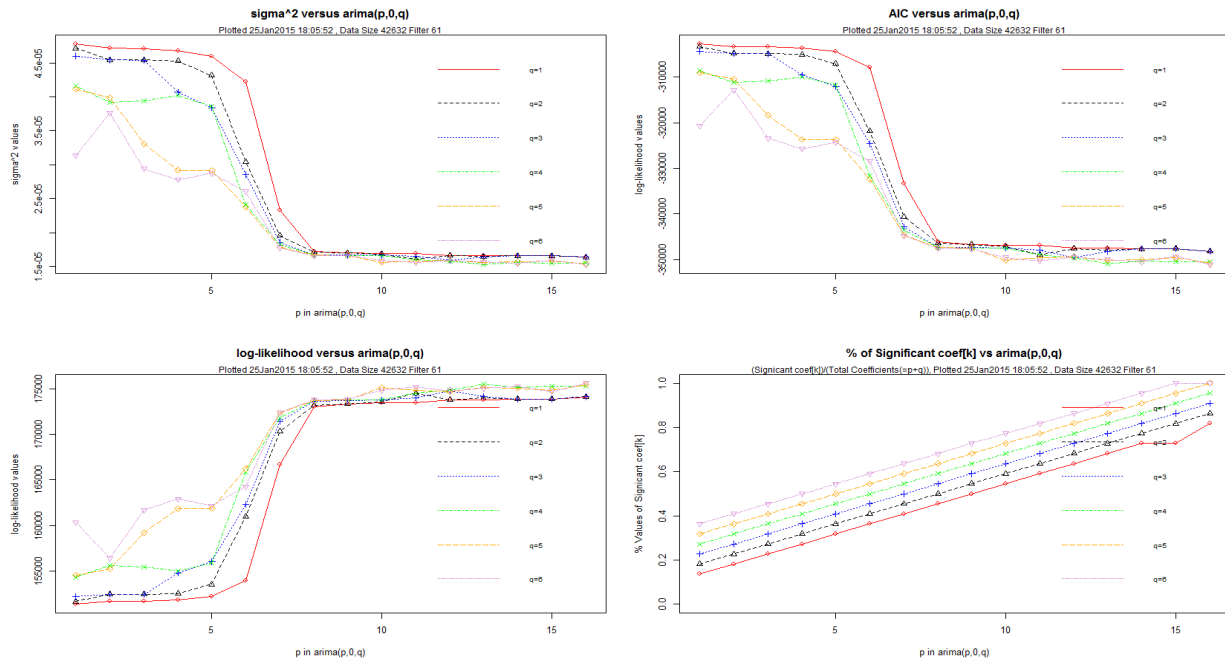


Figure 62: The effect of Filter 61. ARIMA(p,0,q) model scan of AR1-16 and MA1-6 for four flights of the first order differenced Variable Geometry (VG) data.

III. METHODOLOGY

Overview

The purpose of this chapter is to demonstrate that the calculated engine compressor stall model can be an effective predictive maintenance indicator for the data used from the historical General Electric TF34-100 engine data repository. This study demonstrates that the probability model developed is adequate according to the theoretical model adequacy checks.

Test Subjects

Once again, the engine sensor data is a Time Series data. This raw data is collected from the RTED data recorders of the General Electric TF34-100 engines, which are installed throughout the A-10 fleet. The flight sensor raw data is used to calculate the VG data and first order differenced VG data, which then is fitted with an $\text{arima}(p,0,q)$ model. From the results of the ARIMA modeling, a probability model is then fitted to determine the likelihood of a compressor stall occurring.

To ensure that a high fidelity is maintained in the study, much effort went into the elimination or reduction of the impact from two known variance issues. First is the sensor variance. This is an inherent variance associated with the nature of real world engineering data. Secondly is the modeling variance. This variance is related to the adequacy of the proposed model.

The following strategic steps were designed within the research:

First is to collect sample data and conduct a primary investigation. This includes making observations using existing tools to observe which events will be detected after flight. The data needs to be filtered and manipulated. This is to filter out or reduce the extent of sensor failure,

sensor normal noise, and sensor abrupt noise. Next is developing a model by trial and error.

Here an adequate ARIMA model is identified with its necessary R scripts such as:

$$X_t = \delta + AR_1X_{t-1} + AR_2X_{t-2} + \dots + AR_pX_{t-p} + A_t - MA_1A_{t-1} - MA_2A_{t-2} - \dots - MA_qA_{t-q}$$

Where X_t is the VG value at the time t , AR_p is the AutoRegression (AR) coefficient, and MA_q is the Moving Average (MA) coefficient. This model is denoted as arima(p,0,q).

Next is the collection of a large population data and to develop the associated model.

This includes; identify the model variables and select a proper ARIMA model, estimate the ARIMA parameters, check for model adequacy, and fit a linear regression probability model for compressor stall symptom. This model is given by:

$$p(EngineStall|N_{flights}) = \beta_0 + \sum_{i=1}^p \beta_i * AR_i + \sum_{j=1}^q \beta_{p+j} * MA_j + \epsilon$$

Where $N_{flights}$ indicates number of flights data used, currently, $N = 1, 2, \dots, 7$. AR_i is the AR coefficients ($i = 1, 2, \dots, p$) and MA_j is the MA coefficients ($j = 1, 2, \dots, q$).

This model of compressor stall fault event is based on the data as described below:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} \sim \begin{pmatrix} AR_{11} & AR_{21} & \dots & AR_{p1} & MA_{11} & MA_{21} & \dots & MA_{q1} \\ AR_{12} & AR_{22} & \dots & AR_{p2} & MA_{12} & MA_{22} & \dots & MA_{q2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ AR_{1n} & AR_{2n} & \dots & AR_{pn} & MA_{1n} & MA_{2n} & \dots & MA_{qn} \end{pmatrix}$$

Where the response variable (independent variable) (y_1, y_2, \dots, y_m) indicates compressor stall fault event did occur for engine #1 to engine #m, and (y_{m+1}, \dots, y_n) indicates no stall fault event occurred for engine #(m+1) to engine #n; AR_{ik} ($i = 1, 2, \dots, p$; $k = 1, 2, \dots, n$) and MA_{jk} ($j = 1, 2, \dots, q$, $k = 1, 2, \dots, n$) are the repressor variables (explanatory variables) which are AR

coefficients and MA coefficients of the fitted $\text{arima}(p,0,q)$ model from the each individual set of the first order differenced VG data.

Currently, the $m = 6$, $n = 28$. That is, among the 14 aircraft studied (equipped with 28 engines), 6 engines have compressor stall fault events out of the 28 total engines.

Summary

By fitting an $\text{arima}(p,0,q)$ model from the RTED data for each individual engine of the A-10 fleet, a Probability Model for Compressor Stall Fault Event has been established, which can predict the probability of the compressor stall to that particular engine occurring during the next flight. If the probability is high, an engine preventive maintenance will be recommended, thus avoiding the potential costs resulting from the compressor stall damage, and the associated pilot safety issues.

IV. ANALYSIS AND RESULTS

Results of $\text{arima}(p,0,q)$ Model

The Model Adequacy Check for $\text{arima}(p,0,q)$ model of a first order differenced VG data shows very good results (Refer to Figure 29 - Figure 37). The Normal Probability Plot also illustrates a very good fitness for the $(-2,2)$ Normal Score range of data. (The observed tails may be related to the existing data during take-off period and during the landing period.) Two major steps have been taken to ensure the $\text{arima}(p,0,q)$ adequacy.

First, an R script is developed to check the 4 performance parameters of a to-be-specified $\text{arima}(p,0,q)$ over various combined-flight-data . The performance parameters are: Sigma^2 , $\text{Log} - \text{Likelihood}$, AIC , and Percentage of Significant Coefficients (PSC). (See Figure 22, Figure 23, Figure 24, Figure 25, Figure 26, Figure 27, and Figure 28.)

Second, another R script is developed to check the residuals of a specified $\text{arima}(p,0,q)$ model. The following plots are checked: Ljung-Box test p-value vs Lag; Residuals Sequence, ACF of Residuals, Histogram of Residuals, Normal Probability Plot of Residuals, and Standardized Residuals.

The results of the examinations mentioned above are very promising. No major obstruction was observed. During the process, $\text{arima}(20,0,6)$ was selected as the desired ARIMA model to be applied. The imposed time-constraints lead to the selection of a total of 28 sample engines, which turned out to be smaller than the desired sample size.

Results of Linear Regression Model (LRM)

The Model Adequacy Check for the Probability Model of the Compressor Stall Fault Event is accomplished in two parts.

First, R built-in functions for linear regression `plot(lm{stats})` are used, which compare the 4 plots as: Residuals vs Fitted Values, Scale –Location, Normal Probability, and Residuals vs Leverage (Cook’s Distance). For examples of these refer to: Figure 38, Figure 40, Figure 42, Figure 44, Figure 46, Figure 48, Figure 50, and Figure 52.

Second, the Response vs Estimated Response plots (y vs \hat{y} plot) are checked as shown below as an intuitive visualization presentation. Blue dots with dashed lines illustrate the original responses which are detected by ASIST. The black points with solid lines illustrate the estimated response by a fitted LRM. The order of `arma(p,0,q)` used and the resulting modeling parameters are also shown on each of the exemplary plots. Refer to Figure 63, Figure 64, and Figure 65.

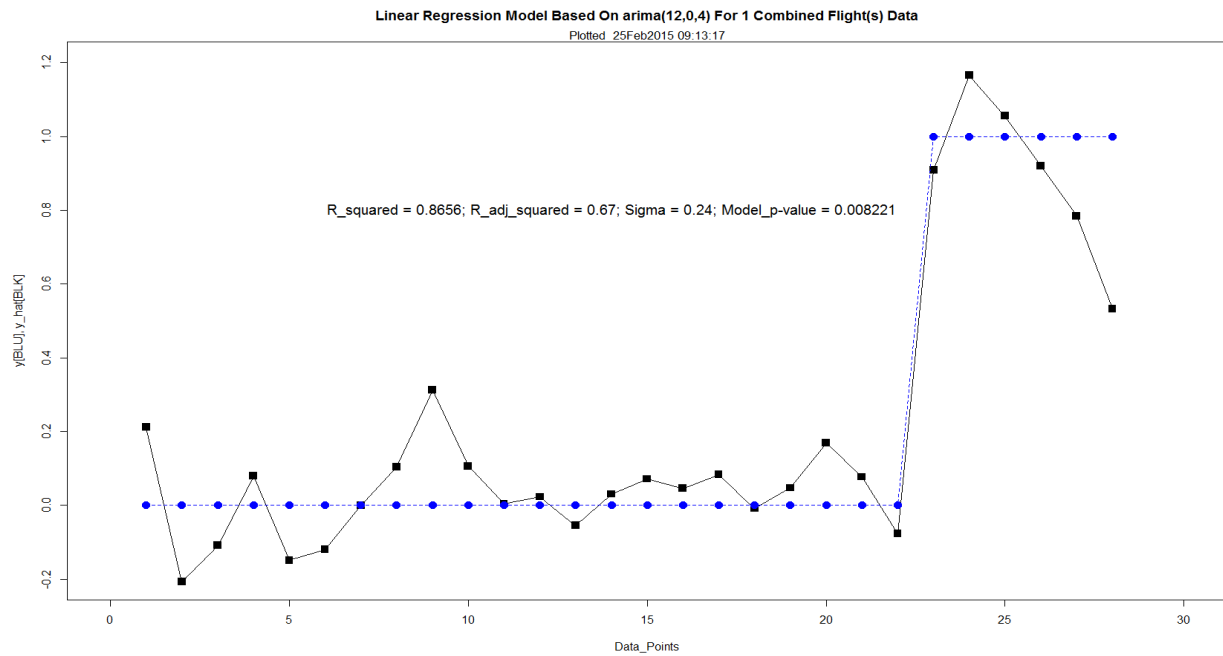


Figure 63: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(12,0,4) for one-Flight-Data.

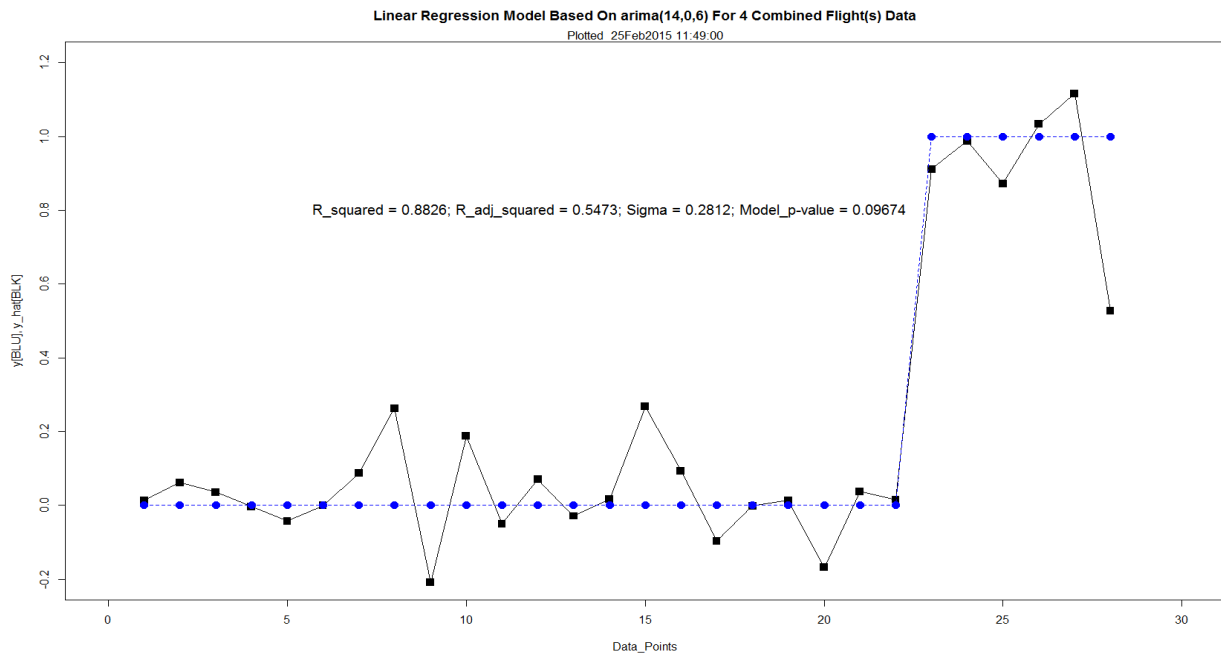


Figure 64: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(14,0,6) for four-Combined-Flight-Data.

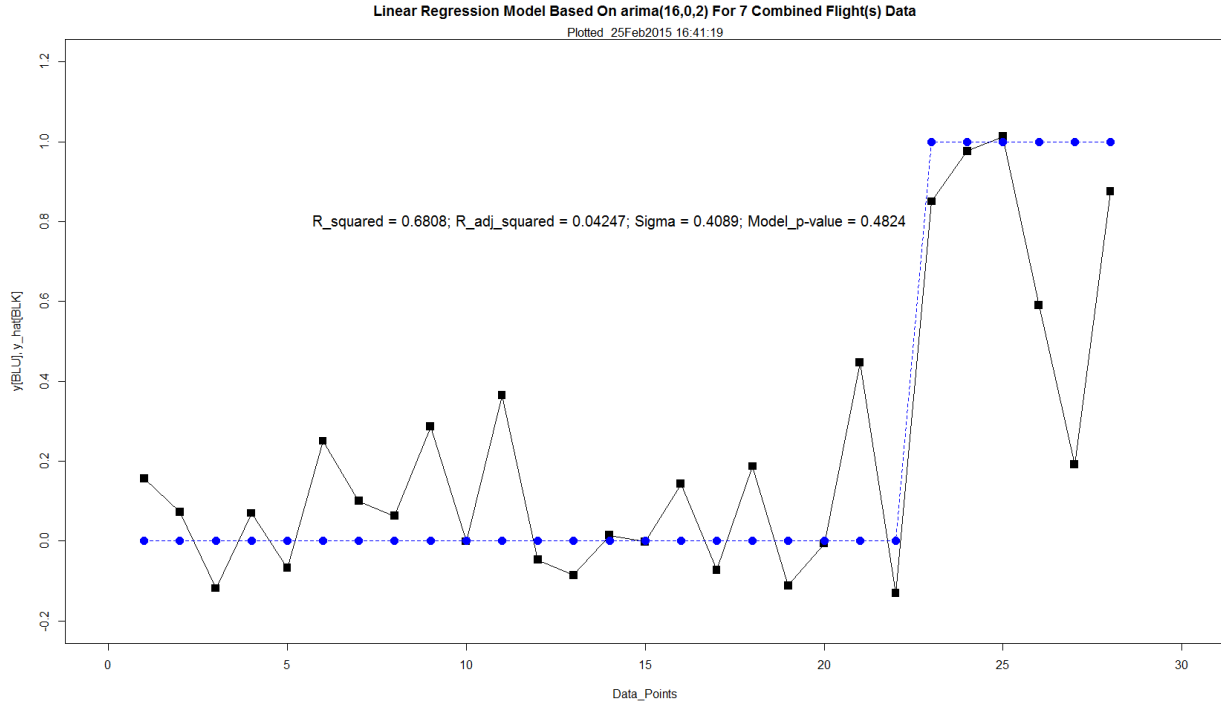


Figure 65: Check the goodness of Fit of proposed Probability Model for Compressor Stall Fault Event. This plot is a Linear Regression Model (LRM) based on arima(16,0,2) for seven-Combined-Flight-Data.

These plots demonstrate strong evidence that when more flight data is used to fit an arima(p,0,q) model, the higher order of an AR would lead to a better LRM. This agrees with the findings of the PACF analysis mentioned in the previous sections. (See Figure 13, Figure 14, and Figure 15.) Utilizing an arima(20,0,6) model indicates a greater LRM, but current sample size impositions have artificially limited that option. Mathematically, when considering the equation

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

Where its number of repressor is a combination of terms (p+q) in arima(p,0,q), it becomes clear that a larger sample size would lead to better R_{adj}^2 in LRM.

For example, if $R^2 = 0.80$ and $R_{adj}^2 = 0.75$ are the desired objectives and the intent is to use the arima(20,0,6) model, then the sampling size of $n = 131$ would be required.

Another example, if $R^2 = 0.85$ and $R_{adj}^2 = 0.70$ are the desired objectives and the intent is to use the arima(12,0,4) model, then the sampling size of $n = 49$ would be required.

Investigative Questions Answered

How to sample the flight sensor data?

The Nellis downloaded data and additional flight data from the JRAMS website were obtained for review. A systematic review of the flight data was implemented following execution of the fleet wide A-10 Engine Data Record Program.

How to derive a response indicator (independent variable) in order to establish a regression model?

The existing ASIST software is used to identify Compressor Stall fault events for all the flight data files from the 28 engines of 14 aircrafts.

Is the ARIMA model adequate for the sensor data?

The research proves that the PSC (percentage of significant coefficients) of the arima(20,0,6) model is very high, and the Ljung-Box test; the normal probability plot of residuals and ACF of residuals also have shown very good results. Thus it is believed that a good ARIMA model was developed from the first order differenced VG data.

How do we predict the fault event probability by utilizing the ARIMA model?

A choice was made to use the retrieved arima(p,0,q) model coefficients of first order differenced ts(VG) data in order to establish the linear regression probability model. This model demonstrated a very proficient predicting capability.

V. CONCLUSIONS AND RECOMMENDATIONS

Conclusions of Research

First, this study has successfully developed a method to use the proposed ARIMA-LRM method to predict engine compressor stall fault based on existing real time performance data (RTED) for the A-10 TF34-100 engines.

Second, this modeling can now lead to significant maintenance cost savings for the USAF, increase the A-10's reliability and, subsequently, the pilot's safety.

Significance of Research

This is the first time a method has been proposed to predict the compressor stall fault event probability of an individual A-10 TF34-100 turbofan engine based on RTED data.

Recommendations for Future Actions

Establish an Engine Data Analysis team to monitor, track, and analyze A-10 fleet-wide engine data.

Improve the proposed Probability Model for Compressor Stall Fault Event by utilizing all the available data for a larger sample size and a longer time period.

Conduct a comparison of "Cost of Maintenances with Predicting Implemented" to "Cost of Maintenances without Predicting Implemented".

Recommendations for Future Research

Investigate the use of a Kalman Filter for the ability to suppress variance, and a better model.

Use a Fast Fourier Transform to achieve higher computation efficiency. Investigate other perspectives to achieve better models.

: Investigate whether a multivariate approach would yield a better model by adding more sensors into the explanatory variables.

Use Bayesian Statistics to test the trustworthiness of the model.

Study the effects of adding more engine faults into the response variables in an effort to gain a better understanding of the engine system. This would be a more comprehensive undertaking, and the complexity of work would be increased exponentially.

Summary

An ARIMA-LRM method has now been developed to predict engine compressor stall fault based on real time performance data (RTED). The ARIMA-LRM results for 1 flight-data, 3- and 4-combined-flights-data are very good. The ARIMA-LRM results for 2-, 5-, 6- and 7-combined-flights-data are not as desirable. This model can be exploited to achieve a great benefit in cost savings for engine maintenance repairs, longer engine lifetimes, and also in increased aircraft/pilot safety. That is, this method can save many millions of maintenance dollars, increase engine/mission reliability, and ensure a greater safety for A-10 pilots.

BIBLIOGRAPHY

Luca Marinai, Rita Singh, Barry Curnock and Douglas Probert, “Detection and Prediction of the Performance Deterioration of a Turbofan Engine”, Proceedings of the International Gas Turbine Congress 2003 Tokyo, November 2-7, 2003

S. Ganguli, S. Deo and D. Gorinevsky, “Parametric Fault Modeling and Diagnostics of a Turbofan Engine”, IEEE CCA/ISIC/CACSD, Sept 2-4, 2004, Taipei, Taiwan

Northrop Electronics Division, Field Engineering Unit/H564, “Field Engineering A-10 TEMS Software Handbook”, A-10 Turbine Engine Monitor System, for Version 54.1 and 56.1 software, section 3.1.4 VG Schedule.

APPENDIX A. THE FORMULA TO CALCULATE VARIABLE GEOMETRY (VG)

Variable Geometry (VG) Schedule Calculation Formula

$NGC = \frac{NG}{\sqrt{(T2C + 273.15)/288.15}}$		
NGC Range (%)	VG = (°, degree)	Note
37.7 °C ≥ T2C ≥ 23.8 °C		
(56.1798, 69.0730)	$IGV - 0.556883 * NGC + 93.385559$	Formula #1
(69.0730, 73.0337)	$IGV + 1.355821 * NGC - 148.570620$	Formula #2
(73.0337, 77.0056)	$IGV + (0.861517 * NGC - 112.469742) + \{(1.005462) * NGC + (-71.408119)\} * (T2C - 23.888889) / 13.888891$	Formula #3
(77.0056, 78.6517)	$IGV + (0.861517 * NGC - 112.469742) + \{(0.577181) * NGC + (-38.428047)\} * (T2C - 23.888889) / 13.888891$	Formula #4
(78.6517, 82.7191)	$IGV + (2.276908 * NGC - 223.792650) + \{(-0.838210) * NGC + (72.894861)\} * (T2C - 23.888889) / 13.888891$	Formula #5
(82.7191, 85.3933)	$IGV + (2.276908 * NGC - 223.792650) + \{(-1.303533) * NGC + (111.385942)\} * (T2C - 23.888889) / 13.888891$	Formula #6
(85.3933, 93.3933)	$IGV + (0.973375 * NGC - 113.258917) + \{(0.000000) * NGC + (0.852209)\} * (T2C - 23.888889) / 13.888891$	Formula #7
T2C < 23.8 °C		
(56.1798, 69.0730)	$IGV + 0.556883 * NGC - 93.385559$	Formula #8
(69.0730, 73.0337)	$IGV + 1.355821 * NGC - 148.570620$	Formula #9
(73.0337, 78.6517)	$IGV + 0.861517 * NGC - 112.469742$	Formula #10
(78.6517, 85.3933)	$IGV + 2.276908 * NGC - 223.792650$	Formula #11
(85.3933, 93.3933)	$IGV + 0.982 * NGC - 113.258917$	Formula #12
(93.3933, 100%)	$IGV - 21.5$	Formula #13
T2C > 37.7 °C		
(56.1798, 69.0730)	$IGV + 0.556883 * NGC - 93.385559$	Formula #14
(69.0730, 77.0056)	$IGV + 1.866979 * NGC - 183.977861$	Formula #15
(77.0056, 82.7191)	$IGV + 1.438698 * NGC - 150.897789$	Formula #16
(82.7191, 93.3933)	$IGV + 0.973375 * NGC - 112.406708$	Formula #17
(93.3933, 100%)	$IGV - 21.5$	Formula #18

What to do if NGC
< 56.1798

What to do if NGC
> 93.3933

VI TA

Shuxiang ‘Albert’ Li,

Born and raised in a small town, Jiang’an Sichuan, by the Yangtze River bank in the Sichuan Basin, China. Graduated from Nanjing Institute of Posts and Telecommunications in 1984.

Worked for a Fiber-optic semiconductor company in Wuhan Hubei for over 12 years.

Immigrated to California in 1997 to work with multiple Fiber-optic passive components companies. Currently work for the United States Air Force at Hill AFB since 2006. I have a loving wife of 26-years, 22 year old and 14 year old daughters, and an 8 year old son. I enjoy spending time with my family hiking and fishing, and building DIY Hi-Fi amplifiers and speakers.

G. Trevor Jones,

Born in Provo, Utah. Raised in the California high Mojave Desert. Graduated with a Bachelor’s of Science in Electrical Engineering degree from the University of Utah in 1986. Obtained a Master’s in Business Administration degree from California Coast University in 2013. Worked for many defense contractors in the missile industry prior to working on the B-52 simulator at Hill AFB in 2009. Currently work for the A-10 SPO Avionics Group WWAEA. I have a loving wife of 33 years, eight children and eight grandchildren. I enjoy spending time with my family in the outdoors, traveling and while staying at home. I have a passion for searching for lost Spanish mines/legendary treasures in the Wasatch Mountains.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 074-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 18-06-2015		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From – To) March 2014 – June 2015	
TITLE AND SUBTITLE A Method To Predict Compressor Stall In The TF34-100 Turbofan Engine Utilizing Real-Time Performance Data				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Jones, G. Trevor, Mr., Civ, USAF Li, Shuxiang 'Albert', Mr., Civ, USAF				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENY) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENV-MS-15-J-036	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Intentionally left blank				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RHIQ (example)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.					
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.					
14. ABSTRACT The Air Force current operations continue to undergo significant changes compelled by decreasing fiscal appropriations, aging aircraft, and personnel drawdown. The Air Force must effectively improve current maintenance operations in part to deal with these challenges. This study will explore the area of the A-10 aircraft fleet's TF34-100 high-pass turbo-fan engine sensor data to seek its deterioration modelling and prognostics capability. In futurity this will allow for achievement of greater confidence in predicting the compressor stall which leads to engine performance deterioration and a costly repair in maintenance. By utilizing an innovative method to forecast the probability of compressor stall, according to individual engine sensor data which has recently become available, it will be possible to achieve significant benefits in both maintenance planning and mission scheduling (which will greatly reduce the associated costs of maintenance servicing).					
15. SUBJECT TERMS Engine Compressor Stall, Predict, Turbopan, ARIMA, Linear Regression, Real Time					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 100	19a. NAME OF RESPONSIBLE PERSON Maj. Jason Freels, PhD, AFIT/ENV
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (937) 255-3636 x4676 jason.freels@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18